Automatic Soccer Video Summarization

Thesis · July 2008

Hossam M. Zawbaa
Beni Suef University

Mohamed Eldib
Ghent University

Cairo University

Faculty of Computers and Information

Information Technology

Graduation Project

ASVS

2008

Automatic Soccer Video Summarization

Team members:

1- Bassam Sherif Hassan Abou-Zaid
2- Hossam Mohammed Zawbaa Ismail
3- Mohammad Mahmoud Mohammad El-Zahar
4- Mohammed Yehia Mohammed El-Dib

Supervised by:

Dr. Motaz El-Saban

Contents

List of Figures

List of Tables

Abstract

This thesis we present the idea of using cinematic features extracted from the input soccer video for the purpose summarization using an automative and efficient framework. The proposed system consist of several basic stages such as (dominant color detection, shot-boundary detection, shot classification and replay detection) which they are considered as a low level soccer video processing algorithms combined with some higher level processing algorithms such as event detection.

The output of our system will produces three types of summaries based on the user preferences selections : 1) Goal events, 2) Attack events and 3)Other events (fouls, injuries, offside). We applied more than 10 hours captured at different championships and conditions to test the efficiency and effectiveness of the proposed system.

Acknowledgement

We would like to express our sincere gratitude to Dr. Motaz El-Saban for letting us work on the exciting topic of sports video processing, giving us so much freedom to explore and investigate new areas of video processing, providing invaluable personal and professional guidance, being accessible all the time, and always being ready to answer our so many questions. He has been and continues to be a source of inspiration for us. He always encourages us to produce better results and to be more proactive.

We have found Dr.Motaz to be extremely honest while giving information to us. We have always been impressed of his high-level of motivation and goal oriented. He has a depth of Information Science knowledge, especially in computer vision, pattern recognition and image processing which directed us to the best way.

[1]

Introduction

This chapter introduces our motivation, challenge, methodology, the problem definition, scope of work and finally the system overview

1.1 Motivation

In this age of universal the need of computer-based technology is very essential in many application which require many computations that must be fast and accurate as possible as we can. Two trends have come together to make the topic of this book of vital interest. First, the explosive growth in computer systems has increased the dependence of both organizations and individuals on the information processing using these systems. This, in turn, has led to a heightened awareness of the need to make this processing using a developed application. Second the huge and the continuous explosion of multimedia data especially video in the current era emphasizes great needs of video database to make indexing and querying to extract the relevant information needed using computer-based application. This is an essential issue to present which attracted great research attentions to make detection of semantic events from several types of videos.

As an example of computer vision disciplines which involved the usage of video processing techniques to perform different operations is the sport video such as (soccer, tennis, basketball … etc). Sports attract many people around the world especially soccer, in the past, people were watching their local league, but now with the evolution of communications (Satellite, Internet … etc) they can watch more than one league at the same time around the whole world.

We have two propellants which motivated us to develop an automated system for soccer match summarization. First, for individuals; most people cannot watch all matches which are played at the same time within different time zones because of lack of time. Second, for professionals; coaches need to view the highlighted events to truly developing plans and to evaluate their team players. From this point we concluded the importance of our proposed program to put a solution for the mentioned problem.

1.2 Problem Definition

Summarization process is an essential part in several applications such as (Information retrieval, video retrieval, speech retrieval … etc), we need to retrieve only an important data from a whole one. This field is undergoing rapid change, as computers are now prevalent in virtually every application, from games for children through the most sophisticated planning tools for governments and multinational firms.

Suppose that we have a soccer match and we need to summarize it by using a computer-based application, our concern here is to extract the most exciting events in the soccer game such as goal and goal attempts using our proposed application afterwards output those events into summarized video.

1.3 Scope of Work

There are many kinds of sports such as basketball, tennis, baseball, hockey, and soccer. We choose to work on soccer game, because it is a popular sport around the world. It has many challenges such as different transitions, hard cuts, and long duration. We worked in three championships Africa Championships League 2008, European Championships League 2008 and Euro 2008 as a starting point of datasets and we can increase the number of these championships in the future, because there are different approaches to differentiate events from one championship to another and the various approaches considered in our proposed system dedicated for these championships.

1.4 Methodology

We would like to represent a general idea of the methodology of the project. The project team was working in a group of two teams, each team has two members, each team selected one or more of the techniques implemented in each topic of our work. And then collect all of these techniques together.

First of all, a thorough search was a help to gather as much as we can of information and recently publish techniques about the target subject: Automatic Soccer Summarization Video. Various papers about different techniques were found, some of them were recently published and others were old published. The selected techniques in Automatic Soccer Summarization Video were according to their ease of implementation and as recently of the technique as possible. However, during implementation in programming phase; we had to adopt slight changes due to our requirements.

1.5 Project overview

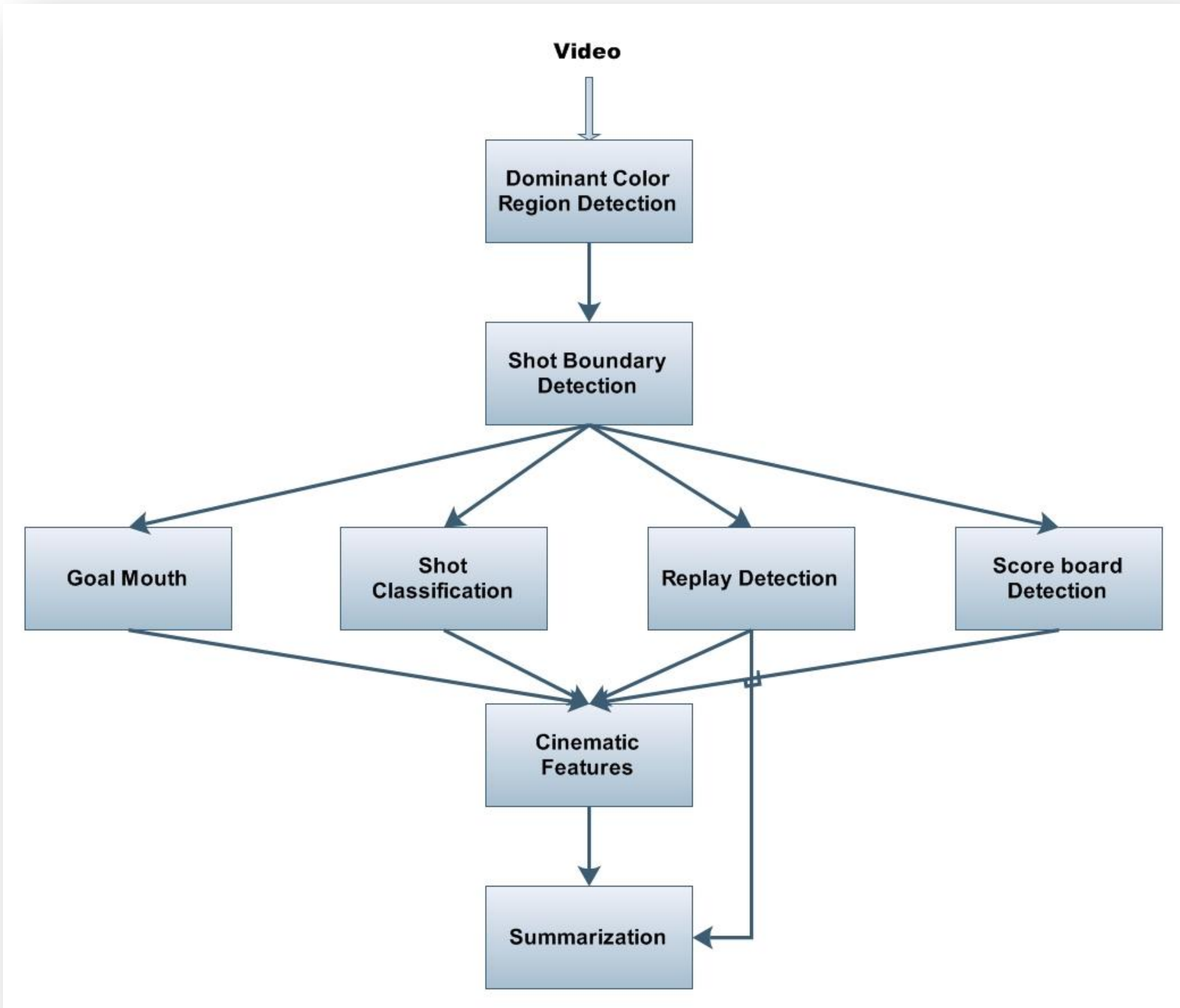


Figure (1.1): Flowchart of the proposed system

The rest of the chapters will talk about the applied techniques we implemented according to the proposed flowchart in Figure (1.1) as the following:

In **chapter 2** we discussed several existing systems in soccer. We mentioned the different approaches of analysis in soccer summarization, and for who those programs are.

In **chapter 3,** this stage is important task during soccer summarization, which the rest of tasks are depending on it. When the accuracy in this task is high, the final output will be more accurate. The idea behind the dominant color detection is to find the most dominated color which it's the green color of the soccer pitch. The dominant color detection is considered as the fundamental process that must be done before shot-boundary detection and shot classification.

There are several techniques of dominant color implementation such as CrCb, rg, HSI, L*a*b*, and RGB, we will talk about HSI and RGB techniques and what are the modifications we applied to fit in our program.

In **chapter 4,** there are several different cameras distributed among the pitch, each one cover a specific view, during the game, there are transitions between those cameras which will produce many views such as, camera 1 focused on the penalty area, while camera 2 focused on corner area, the transition from camera 1 to camera 2 will produce a new scene, our involved task is to extract these scenes.

We show the applied used techniques in Shot-Boundary Detection such as the difference in color histogram similarity, the difference between dominant colored pixel ratios of two frames and the dominant color pixel ratio in a frame. We represent a frame by its color histogram, which is defined in the HSI space.

In **chapter 5,** in this stage, we classify the shots according to the different camera views such as, close-up view, long view, medium view and Audience (out-of-field) view.

In **chapter 6**, we mention the importance of detect the replay shots .Exciting events are often replayed immediately after they occur. In this chapter you will learn methods used for detecting replay segments to localize semantically important events and then further summarize soccer videos.

In **chapter 7,** in this stage, we will discuss the cinematic features such as (score board detection, goal mouth detection and other features which are discussed before in separated chapters like slow motion replay, logo based replay and shot classification) and these features are used in event detection stage.

In **chapter 8,** in this stage we can use all the cinematic features in a sequence manner, that are used as an indicator of the occurrence of important event such as (goal, goal attempts shots, red/yellow card, fouls and offside). Therefore the summarization process which we aimed to be recognized by a combination of these events, for example the summarized segment may contains of only the goal shots, goal attempts or penalty shots that can be described as important events.

[2]

Related works

This chapter introduces the existing works applied in the area of soccer sport including the difference between soccer analysis and summarization as the main topic.

2.1 Soccer Analysis Vs Soccer Summarization

We are going to mention the difference between soccer analysis and soccer summarization process which are used to present the concept of technology in sports, especially soccer.

2.1.1 Soccer Analysis

"Basic game systems simply analyze the match cassette, allowing basic data input which provides statistical information such as the number of passes and tackles. Depending on the system, this can be done in real-time. *Video montage* software allows coaches to *chop up* and *put together* a digitized video of the match (post match only) to select, extract and visualize the information they want. This can be linked to what is known as a digital time code where specific match events such as headers, shots or passes are inputted (can be done in real-time or post-match) at the precise moment in time where they happened. These events can then be directly accessed and visualized at the click of a button. This avoids having to search through the film using a video recorder which is extremely time consuming" [8].

"Digital cameras have become cheaper and better in image quality. This type of camera can be linked directly to a computer and the film analyzed straightaway, saving time (avoids the digitization of a video cassette produced from non-digital type cameras)" [8].

The limitations of this type of software are the speed and accuracy of data can input. Equipment such as touch sensitive pads and voice recognition has made data input quicker and more accurate. The latter is extremely useful as it allows a match analyst to permanently view the game and input the data without being distracted. However, pitch positions of player actions are determined by simply clicking on a pitch representing the playing area, leading to inaccurate data. This lack of accuracy has led to the development of systems which automatically calculate player positions and movements.

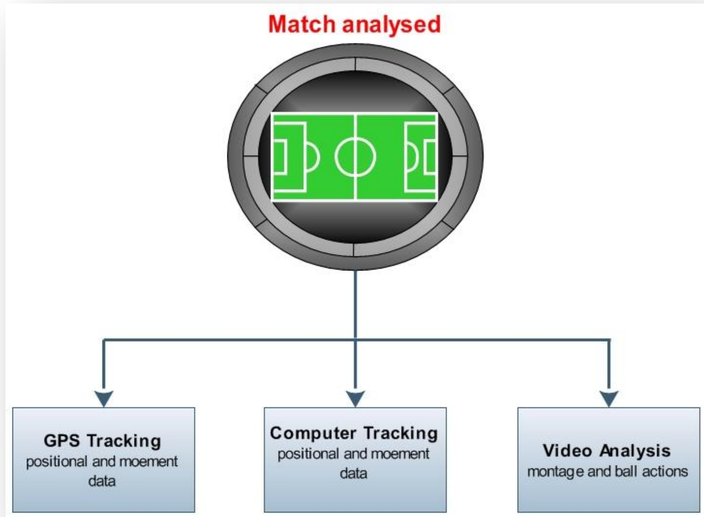


Figure (2.1): The 3 major types of modern computerized match analysis systems – Real-time & post match

"As seen in Figure (2.1) both GPS & Computerized tracking stems may need to be complemented using video montage and that the ball actions such as passes, tackles, headers etc have to be inputted separately. The input of ball actions can be undertaken in real-time but video montage is post-match" [8].

> *"Computerized* systems which *automatically calculate the positions* and *track the movements* of players using camera, video and computer technology are becoming more commonplace. These systems require the installation of several cameras to cover the whole pitch so that every player is always captured on film, whatever their position on the pitch and the moment in time. Using complex trigonometry, mathematical algorithms and digital video/image processing techniques, player positions and movements can be calculated. Due to the positions of the cameras and the difficulty in tracking many moving objects using such complex techniques, doubts may persist on the accuracy of the data produced. As well, the possibility of real-time analysis is extremely limited" [8].

Global Positioning Systems or GPS could well be the future in the computerized analysis of sport. GPS are used by boats to determine their exact positions while portable systems are available and are useful for mountaineering expeditions etc. If GPS can be implemented in soccer training and competition, this would mean an analysis of the movements and physical activity profiles (total distances run, n° of sprints...) of every single player on the pitch and at every moment of the action. Linking this to data produced from training devices such as heart rate testers would allow an excellent picture of the overall player effort.

GPS have the advantage of being extremely accurate and allowing real-time analysis. This means that precise, objective data is rapidly produced and can be given to the coach for immediate evaluation. The disadvantage of such technology being that portable GPS systems are at present, fragile, relatively bulky and expensive. Due to the high impacts often seen in soccer, the microchip transmitters may break unless they are robust or well protected. However, by the time this article is published and with daily technological advances, the latest GPS may now be the size of stamp!

“The data produced from the above mentioned systems must somehow be stored and presented. Modern *databases* allow data to be stored, retrieved, compared and presented in a quick and efficient manner. Player tracking systems allow the whole game to be reconstructed in 2D/3D. This means that the coach can visualize the recreated movements of his players and when linked to the actual video of the game, this offers a powerful means of evaluating and understanding tactical, technical and physical performance” [8].

These performance records can allow detailed comparisons of the player's abilities for team selection and development. For example, analysis may lead to the choice of a player’s most effective playing position or detect a particular weakness in his game and calculate the best possible training program to improve performance

2.1.2 Soccer Summarization

Before talking about soccer summarization process, we need first to present the techniques of sport video processing. Afterwards we will discuss several concepts which are related to sport video especially soccer.

The processing applied into sport video includes different categories of information according to the selected feature that can be extracted such as (visual, audio and text) which can be combined together to extract semantic information (event). These events can be used with different approaches such as (summarization, indexing and retrieval) [1].

Indeed visual information can be considered as the most important feature compared with the mentioned ones, that’s because the visualization of such information has less fault recognition that the other. For example, audio feature may affected by external factors due to speaker characteristics and compression schema [1].

In order to truly detection the high-level events in sport video, we need to extract some low-level features such as (color, texture, shape and motion). In addition cinematic feature such as (shot-types, shot-length and replays) introduce another factor that we need when we are working on sport video to produce summarized events [1].

When we are talking about soccer game, we can refer to a continuous sports which mean that if there is an existence of such a break during the match, it can an indicator of the occurrence of important event such as (goal, penalty shot and red/yellow card). Therefore the summarization process which we aimed for can be recognized by a combination of these events, for example the summarized segment may contains only the goal shots, goal attempts or penalty shots that can be described as important events [1].

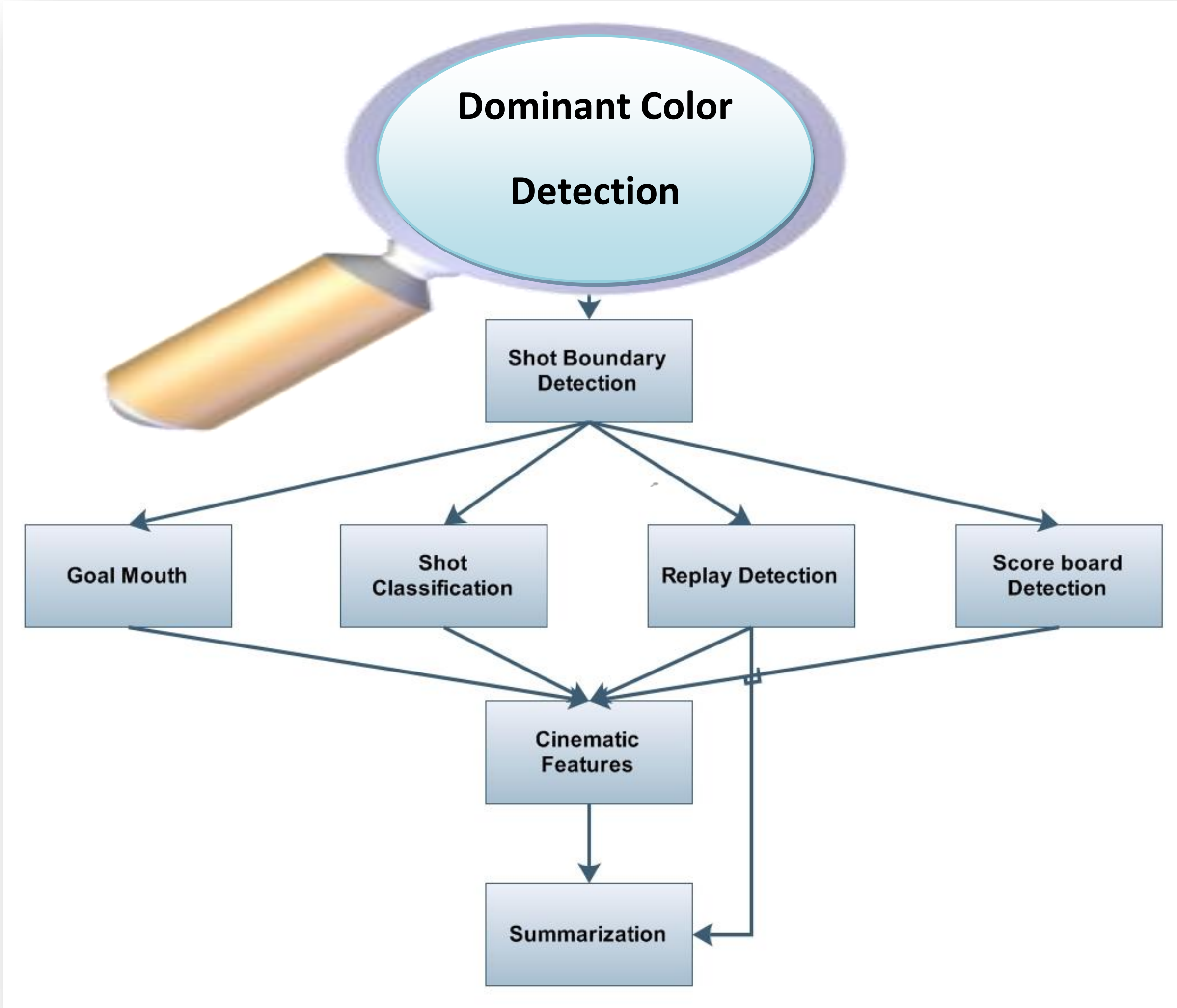
In this approach, we are going to highlight the most important events such as (Goals, Goal attempts), facilitate the process of automatic match, save the viewer’s time, and introduce the technology of computer-based summarization into sports field.

This is dedicated to professionals to enable sport satellite channels automatically highlight the important events in a match. These help professionals to comment on them and help trainers to analyze their team performance, and to interested users to provide the most important events of an entire match under constraint of limited viewing time.

[3]

Dominant Color Detection

This chapter introduces the concept of dominant color detection, the method in which the problem is solved and comparison between the different applied methods.



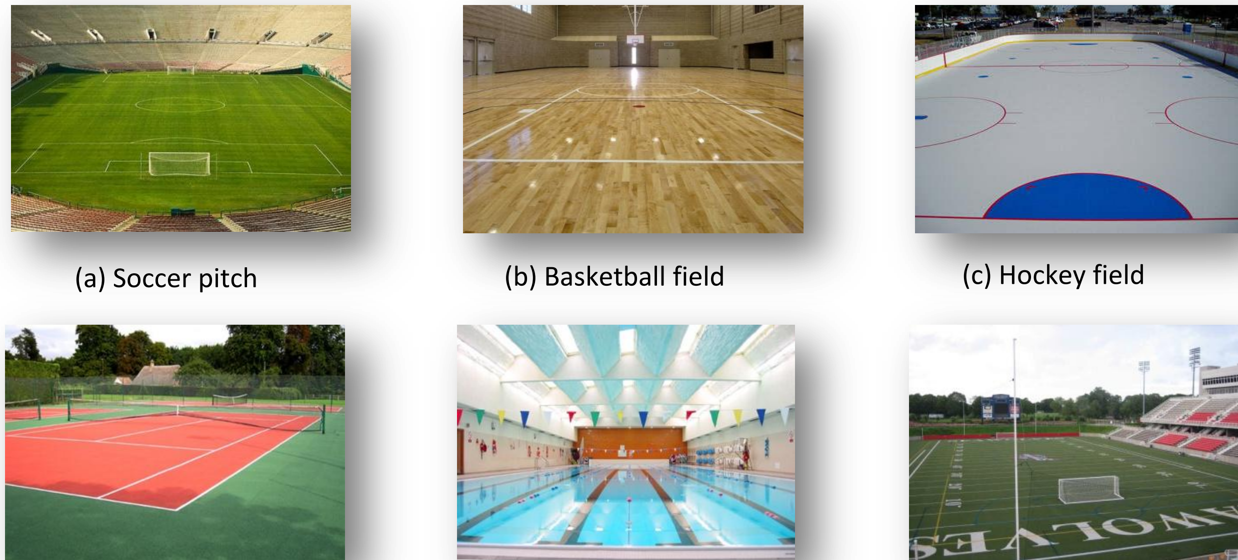
3.1 Introduction

Before starting to identify the basic concepts and techniques used to determine the dominant color region, let us first introduce what is the idea behind the detection of a dominant color and how we can recognize and mark the desired color as the dominated one from other colors.

What is meant by dominant color?

The dominant color is the color which is filling most of the given area, and it is different between various fields. For example, in figure (3.1) we can clearly see that the dominant color differ from one sport to another green in Soccer and Football, brown in

Basketball, Orange in Tennis, white in Hockey and blue in Swimming. We are concerning only with the soccer pitch which is the green color.



(a) Soccer pitch (b) Basketball field (c) Hockey field

(d) Tennis Court (e) Swimming pool (f) Football field

Figure (3.1): Different dominant color fields.

What is the idea behind the detection of the dominant color region?

The field of any sport must contain a unique dominant color which can be used to differentiate among various sports by detecting this field. This stage is a very important one in our proposed system and must be done accurately. When we are referring to shot-classification introduces in the next chapters, we must mention the importance of dominant color that is useful to make this classification. Furthermore, the high-level algorithms such as (object detection and tracking) must apply the dominant color detection operation first [1].

We have different approaches in determining the occurrence of the dominant color region in sport video. The first approach made assumption which says; at the start of soccer game the dominant color region occur in the center frame. While the second approach work with another assumption; by using color histogram of one or more frames, the peak at the beginning sport video can considered as indicator of the occurrence of the dominant color. Indeed the second approach can act accurately compared with the first one regardless of the camera motions and angles variations that’s because it make a global computations [1].

3.2 Color Spaces

Color segmentation algorithms have high importance to extract various features that depend on the selected color space and must be done truly with fewer errors. For this reason we mention the color spaces types and their conversion functions which are used in our system and discuss their usage in the implemented methods [1].

"Hue family of color spaces, such as HSI and HSV; decouple intensity component from the chrominance information that is represented by hue and saturation values. In addition to that, hue and saturation closely match human perception [2]. Although there are multiple definitions, we use Equations for the conversion of RGB to HSI: Since hue is unstable for low saturation values, hue cannot be reliably used for certain range of color values. This range of color values is referred to as the achromatic region in the HSI space "[1].

The most common color space used is RGB. Color Space is defined as a model for representing color in terms of intensity values, RGB stands for Red-Green-Blue. RGB space reduces light effect, regardless of the current weather or stadium condition, or even the quality of the video. In the other hand, HSI doesn't provide the mentioned facilities [9].

What effect the detection of the dominant color?

Most approaches today suffer from factors which have several negative effects on the image quality such as (lighting, shadows, low resolution… etc). And you can't ignore this problem if your goal is depending on the quality that affect on the degree of visibility of relevant information in an image which is appropriate for a specific image task such as (detection, classification … etc) .

3.3 Methods to detect the dominant color region:

The fundamental concepts and algorithms covered in this book are often based on those used in existing commercial systems. Our aim is to present these concepts and algorithms in some easy way which facilitates the understanding of our proposed system.

Now, we are ready to identify the available methods that are used to detect and recognize the dominant color region which are HSI method and RGB method. In the next subsequent sections we compare between the mentioned methods according to our results occurred from the proposed system.

3.3.1 HSI method

As illustrated in Figure (3.2), the first step is to convert the input video file into its corresponding frames, because our actual work is running on to frames. Then we need to convert the produced frames from RGB to HSI color space by using Equations (3.1)-(3.3). Since hue is unstable for low saturation values and Equations (3.1) - (3.3) are undefined for very low and very high intensity values [2], hue cannot be reliably used for certain range of color values. This range of color values is referred to as the achromatic region in the HSI space [1].

$$H = cos^{-1}\left[\frac{\frac{1}{2}[(R'-G')+(R'-B')]}{[(R'-G')^2+(R'-B')(G'-B')]^{\frac{1}{2}}}\right] \quad \textbf{(3.1)}$$

$$S = 1 - \frac{3}{(R'+G'+B')}[\min(R',G',B')] \quad \textbf{(3.2)}$$

$$I = \frac{1}{3}(R' + G' + B') \quad \textbf{(3.3)}$$

"We adopt the formulation in Equations (3.4) – (3.9) to define the achromatic region [3]. In the equation, IMAX and SMAX are the maximum values intensity and saturation channels respectively" [1]. After that, we scan the image pixels and extract the H, S and I components to get their peak.

$$H\,[i_{min}] \geq K.H\,[i_{peak}] \quad \textbf{(3.4)}$$

$$H\,[i_{min} - 1] < K * H\,[i_{peak}] \quad \textbf{(3.5)}$$

$$H\,[i_{max}] \geq K * H\,[i_{peak}] \quad \textbf{(3.6)}$$

$$H\,[i_{max} + 1] < K * H\,[i_{peak}] \quad \textbf{(3.7)}$$

$$i_{min} \leq i_{peak} \quad \textbf{(3.8)}$$

$$i_{max} \geq i_{peak} \quad \textbf{(3.9)}$$

“The peak index, ipeak; for each one-dimensional average histogram is localized to estimate the mean value of dominant color for the corresponding color component, so the next step is to defining interval boundaries in order to remove noise effects that may result from using a single index, an interval about each histogram peak is defined, where the interval boundaries (imin; imax) correspond to the closest indices to the peak index that have less pixel count than some percentage of the histogram peak, which is given in Equation (3.10). In this implementation, we fixed this percentage value as 30% of the peak count, i.e., K = 0.3 in the corresponding equations, where h refers to the histogram of an arbitrary color component” [1].

$$I > I_{low} = 0.10 * I_{MAX} \qquad \textbf{(3.10)}$$

or

$$I < I_{low} = 0.90 * I_{MAX}$$

or

$$S > S_{low} = 0.10 * SI_{MAX}$$

“After the interval boundaries are determined, the mean color in the detected interval is computed by Equation (3.11) for each color component. Q_{size} is the quantization size and is used to convert an index to a color value” [1].

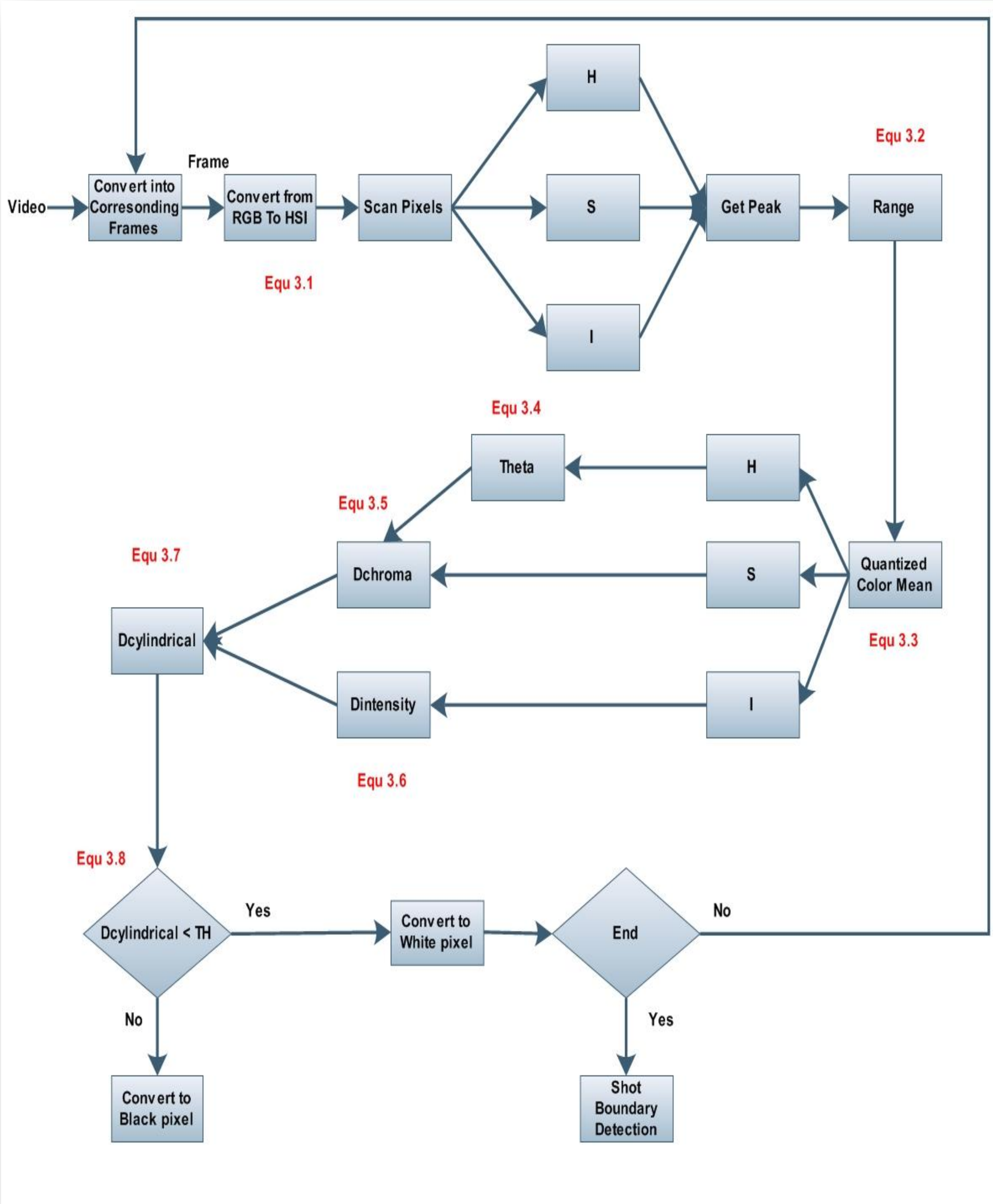


Figure (3.2): Flowchart of HSI dominant color detection method

$$Color\ mean = \frac{\sum_{i=i_{min}}^{i_{max}} H[i]*i}{\sum_{i=i_{min}}^{i_{max}} H[i]} * Q_{size} \quad \textbf{(3.11)}$$

Then we need to compute the (θ) values from the produced H component, by using the Equation (3.12) in order to compute d_{chroma} from the Equation (3.13), at the same time compute $d_{intensity}$ by using Equation (3.14) and after these computations, compute the $d_{Euclidean}$ from the produced parameters by using the equation (3.15). S, and I refer to saturation and intensity, respectively, j and k are the j^{th} and k^{th} pixels, and θ is the minimum absolute difference between the two hue values, i.e., (θ) is limited to [0; π].

$$\theta(j) = \begin{cases} \Delta(j) & if\ \ \Delta(j) \leq 180° \\ 360° - \Delta(j) & otherwise \end{cases} \quad \textbf{(3.12)}$$

$$\Delta(j) = \left|\overline{Hue - Hue_j}\right|$$

$$d_{chroma}\ (j,k) = \sqrt{(S_j)^2 + (S_k)^2 - 2\,S_j\,S_k \cos\ (\theta)} \quad \textbf{(3.13)}$$

$$d_{intensity}\ (j,k) = \left|I_j - I_k\right| \quad \textbf{(3.14)}$$

$$d_{cylindrical}\ \ (j,k) = \sqrt{(d_{intensity}\ (j,k))^2 + (d_{chroma}\ (j,k))^2} \quad \textbf{(3.15)}$$

Finally, we have to calculate the Threshold values in order to specify which pixels are considered as dominant and which are not. This done by summing the pixels values that fall between the defined interval and divide the result by value range between max and min interval, then use the equation (3.16) to make the decision of the dominant pixel.

$$d_{cylindrical}\ \ < \ T_{color} \quad \textbf{(3.16)}$$

At this point, we can recognize the occurrence of the dominant color (green) and convert it to white color as you shown in figure (3.3). We have to repeat the above steps until reaching the end of the video file.

(a) Original Image

(a) Dominant Color region

Figure (3.3): Dominant Color detection result using HSI.

3.3.2 RGB method (The proposed one)

At first, we need to present some factors that motivated us to develop new method in order to determine the dominant color region. The important factor is the processing time in HSI is high, in additional there are several errors occur according to the level of noise that affects on the video file.

So, the mentioned problem enforced us to change our color space to RGB and implement a simple method, through specifying the green region in the pitch. After applying this method, we recognized that, we must specify general color region according to the soccer pitch which is green color as the following:

0<**R**<150, 94<**G**<255, 0<**B**<100

As a result, if any pixel lies between these ranges, it will be a green pixel.

The simplicity of our new algorithm increases the speed of accomplishing the dominant color detection, hence it increment the whole performance of the total output. In additional to that, the detection doesn't affected by the noise applied to the video file.

You can simply follow the steps in Figure (3.4) to implement this method.

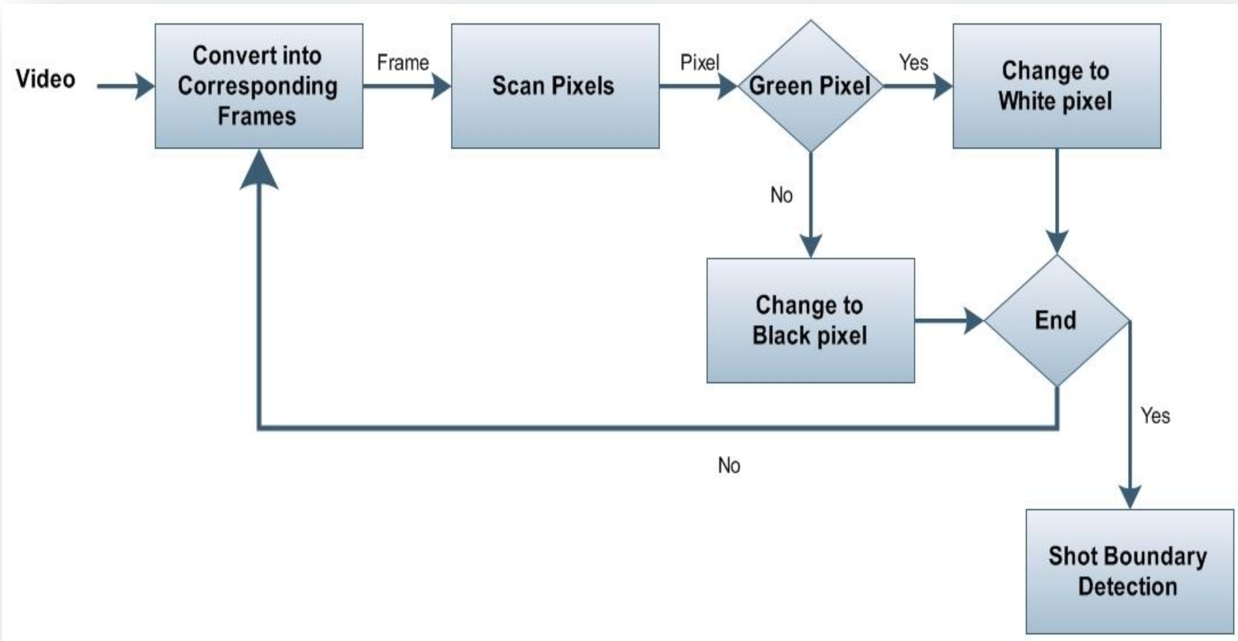


Figure (3.4): Flowchart of the proposed RGB dominant color detection method

Finally, we make the same as we made in HSI, by converting the green field into white color and neglect the other colors as we shown in Figure (3.5).

(a) Original Image

(a) Dominant Color region

Figure (3.5): Dominant Color detection result using RGB.

3.4 Results:

In this section, we would like to present some pictures (frames) resulted from applying the two previous methods in order to clearly compare between them and you will see the huge gap between the two implemented methods.

- Figure (3.6) there are several low level images from a real video file :

    - In HSI method, the common drawbacks is the fault object marking as a green field which mean players are colored with the white color and represents the same color as the field (green) that will lead to wrong results in the next stages.

    - In RGB method, it overcomes this problem. The RGB nature isn't affected by the quality of the image; it can be differentiated easily between objects like players and doesn't mix them up with the green color region.

- Figure (3.7) there are several lighted images from a real video file :

    - In HSI method, the lighting effect on the true recognition of the green color, because it increases intensity of the green color, and turn it to a shining color which consider it as another color.

    - In RGB method, lighting hasn't any effect on the recognition of green color.

- Figure (3.8) there are several shadowed images from a real video file
    - In HSI method, when the shadow effect is the dominated, it considers the shadow as the dominant color; it neglects the actual green color.

    - In RGB method, because of the predefined green color ranges, green color isn't affected by shadow.

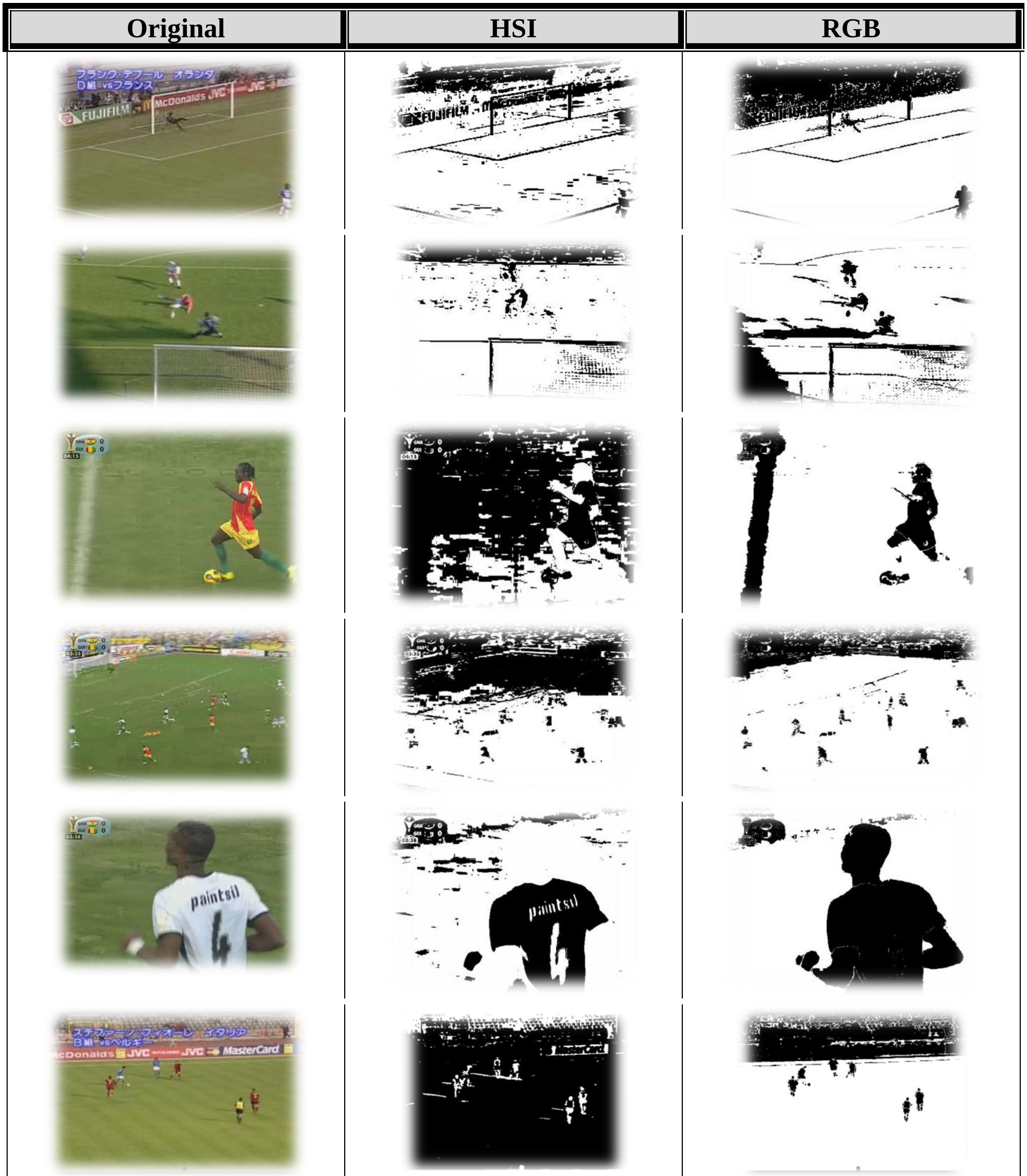


Figure 3.6 : Dominant color detection results with low quality images

Original	HSI	RGB

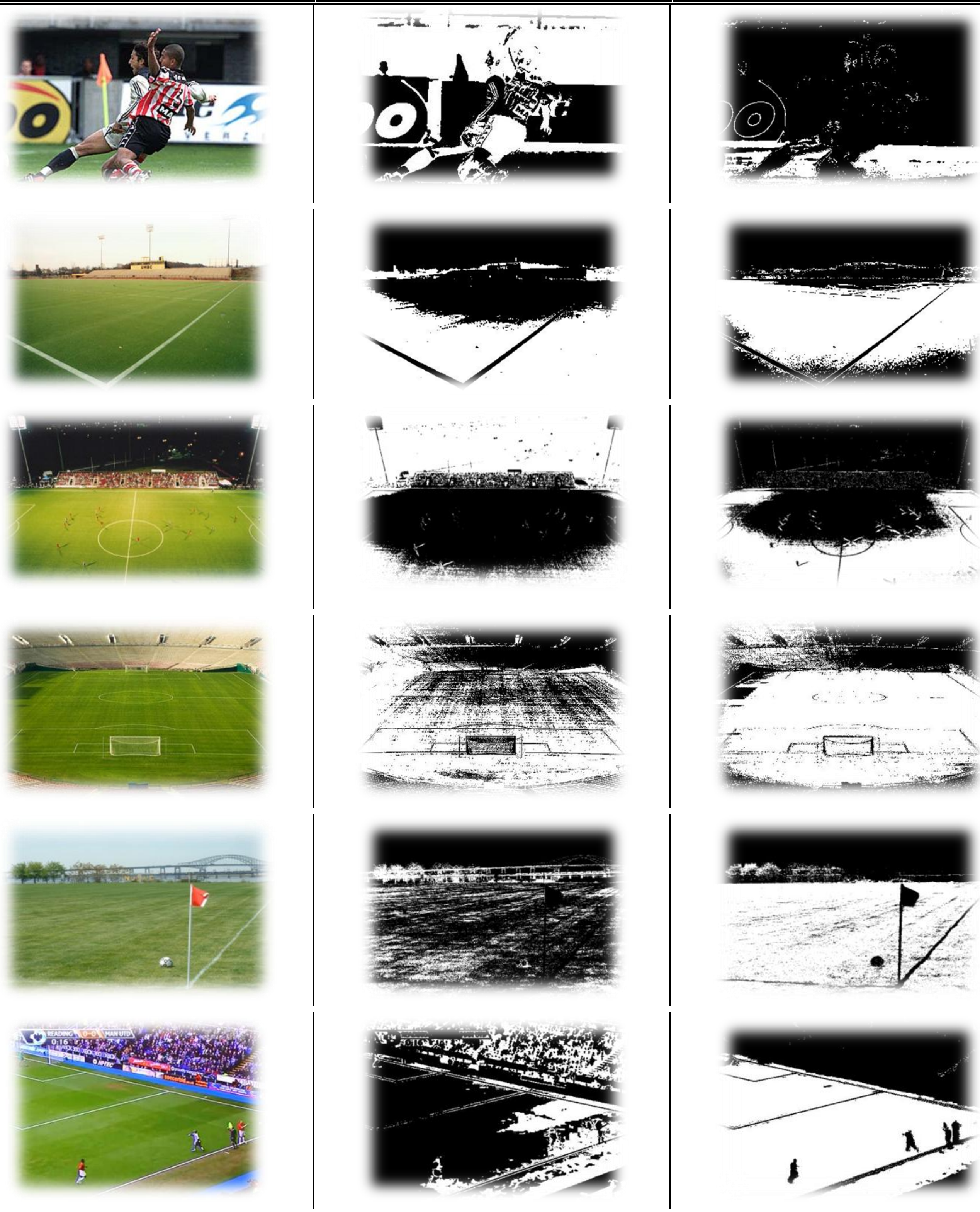


Figure 3.7: Dominant color detection results with high intensity lighting images

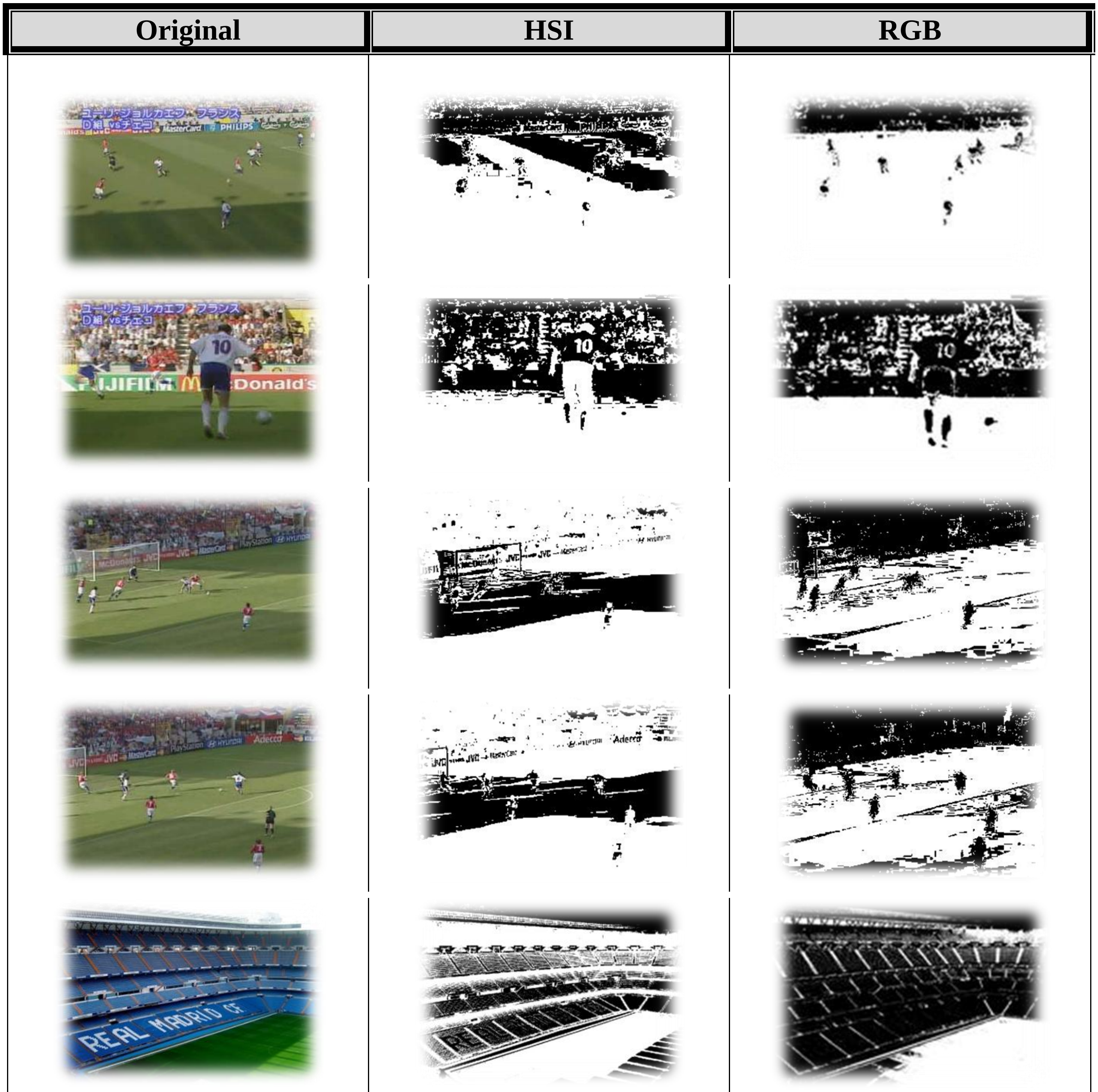


Figure 3.8: Dominant color detection results with shadowed images

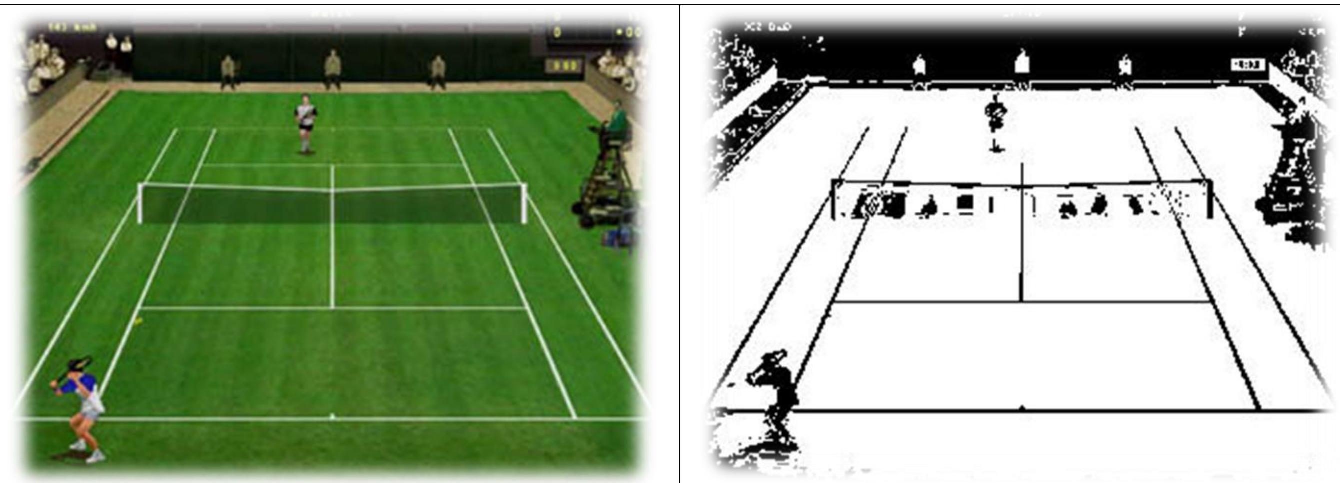
We are going to highlight some factors between RGB and HSI which illustrates the importance of RGB, in the following table depending on our experiments:-

Comparison	RGB	HSI	Figure
Low quality	Good	Bad	3.6
Light effect	Not affected	affected	3.7
Shadow	Not affected	affected	3.8
Camera motion	Good	Excellent	---
Time Consuming	Low	High	---

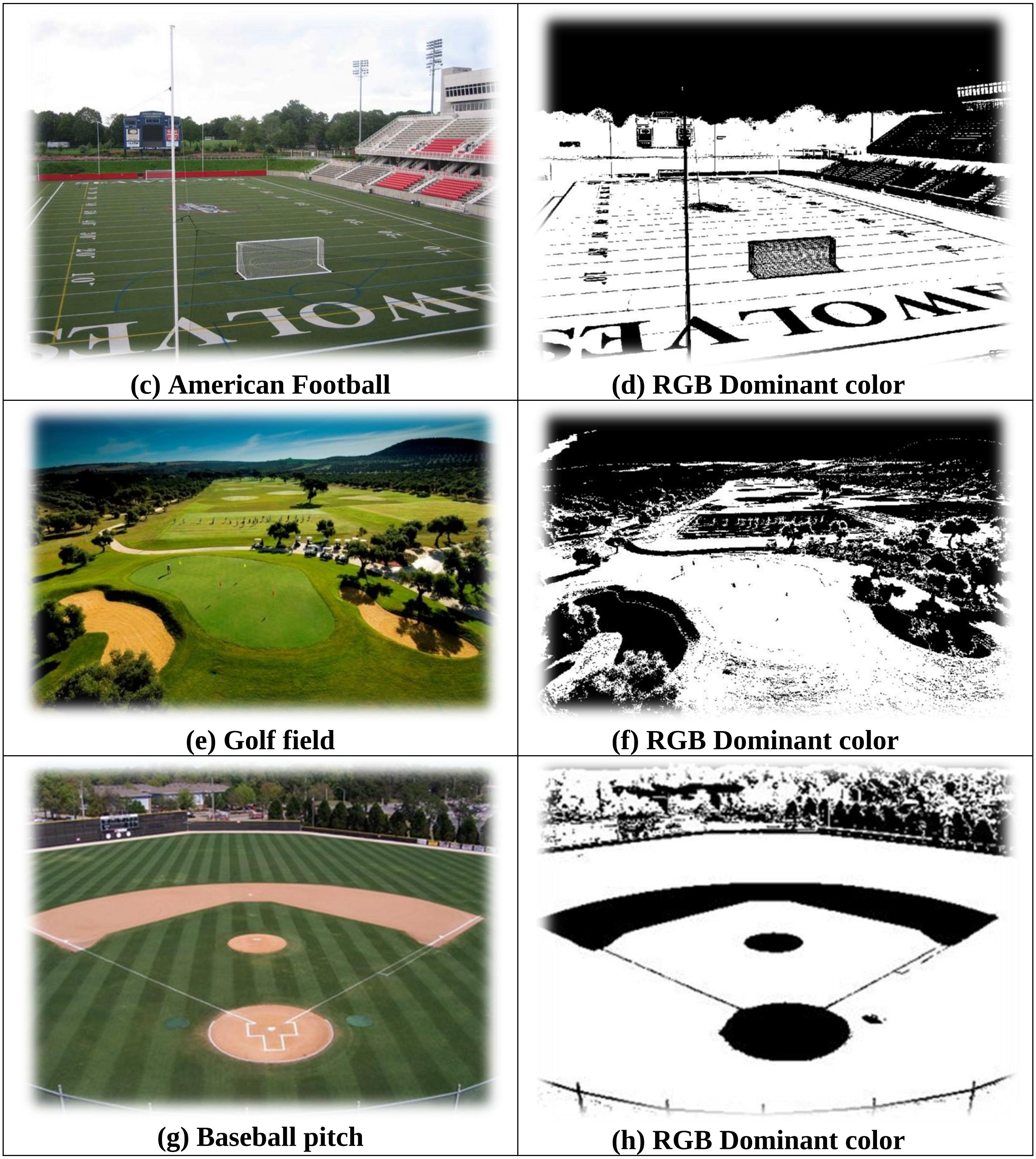
Table (3.1): Comparison between RGB and HSI color spaces in determining the dominant color.

3.5 RGB method with different sports

As we discussed previously that our proposed dominant color method is applied with soccer, we are going to show some working examples of other sports using our proposed dominant color detection method, see Figure (3.9).



(a) Green Tennis court **(b) RGB Dominant color**



(c) American Football | (d) RGB Dominant color

(e) Golf field | (f) RGB Dominant color

(g) Baseball pitch | (h) RGB Dominant color

Figure (3.9): **RGB dominant color detection results applied with other sports**

As a result from the given samples, we can say now that our proposed dominant color region detection is applicable with other sports which have green field such as (Tennis, American football, Golf and Baseball).

3.6 Conclusion:

In this chapter we proposed a new method to detect the dominant color region (green) in the soccer field, which used the RGB color space. So, the domain here is specific to the soccer, but you can detect another domains by modifying the color ranges of the field you're applying this method in it.

This method facilitates the detection process and overcomes the problems produced when applying the HSI method. For example our proposed method (RGB), doesn't affected by the quality of the given video file, also there isn't any effect by different lighting conditions such as weather conditions (sunny, cloudy, raining). Therefore the overall system isn't sensitive to major and permanent variations in the image conditions.

The previous discussion is verifying the powerful of our robust method.

There are some missing parameters in HSI method, we faced while we were working on it such as, the quantization value, we didn’t know what the appropriate value is for it, we concluded its value from a paper, and it wasn’t clearly identified.

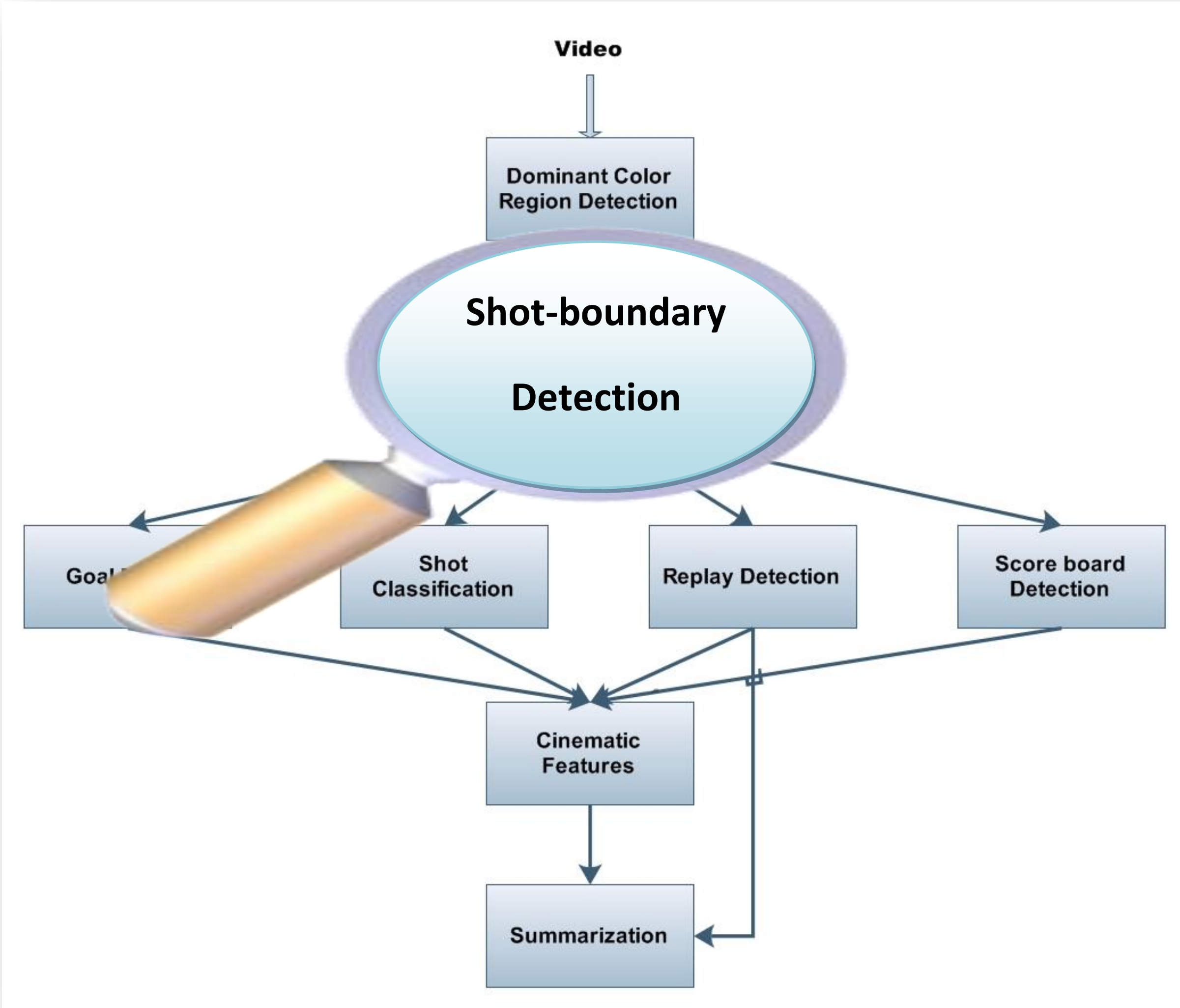
The quantization value for each component:-
Quantization (H) =64, Quantization (S) =64, Quantization (I) =128.

The second missing parameter is, threshold value, which it’s dependent on the video, our proposed solution to this parameter is done by summing the pixels values that fall between the defined interval and divide the result by value range between max and min interval.

[4]

Shot-boundary Detection

This chapter introduces the concept of shot-boundary and the latest applied algorithms to detect the occurrence of new shots. Then by evaluating our works we made some new statistics.

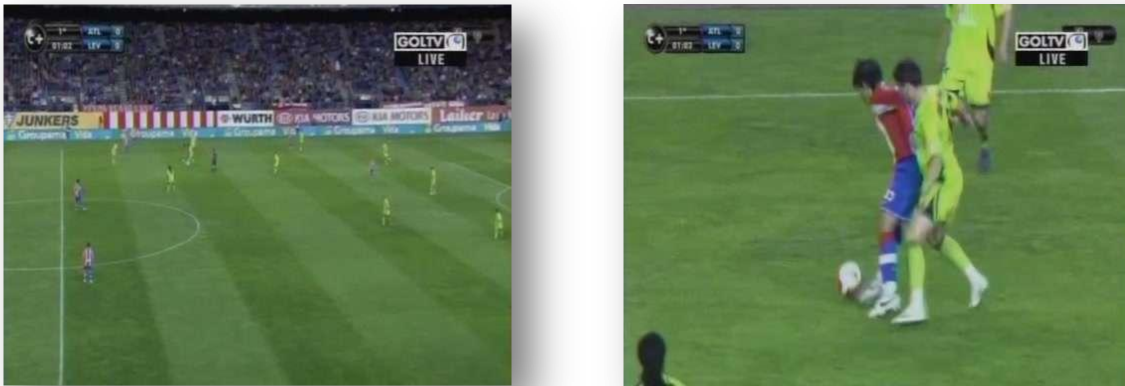


4.1 Introduction:

We will start to define the term of shot which will be used in this chapter, and then we are going to identify the basic concepts and techniques applied in shot- boundary detection.

What is meant by shot?

It is separated view which comes from multiple cameras views positioned at different locations along the pitch; see Figure (4.1) for more details. You can realize that when the changing from camera to another (camera 1 to camera 2) done, the resulted view indicates that it is a new shot and it must be marked as the boundary of the new shot.



(a) Camera 1 (b) Camera 2

Figure (4.1): Example of two different shots

4.1.1 The idea of shot-boundary detection process:

The process of detecting the actual boundary of shots depending on finding the difference between frames by computing frame-to-frame differences in some features and look for a large difference value that should indicate the start of a new shot [1].

The correlation between frames in the same shot is a very important indicator to detect the similarity between them and when there is noticeable change we can conclude the starting of new shot as we mentioned. You must be aware that there are different transitions that may occur during the match, so the next section illustrates this point in more details.

4.1.2 Types of shot transitions

Before discussing the types of transitions occur in soccer we must first mention that the false recognition of new shot resulted from the object and camera motion. We have two types of transition depending on the camera movement and transition; 1) instant (cut) transition and 2) gradual transition. When there is a special editing effects during the match to illustrate the transition we said; gradual transition while the instant transition haven't any effects and it is more accurate than the gradual one [1].

As seen in Figure (4.2), it's an example that illustrates the instant (cut) transition between two shots [(a) to (f)]. While Figure (4.3) show us the second type of shot transition which its gradual transition.

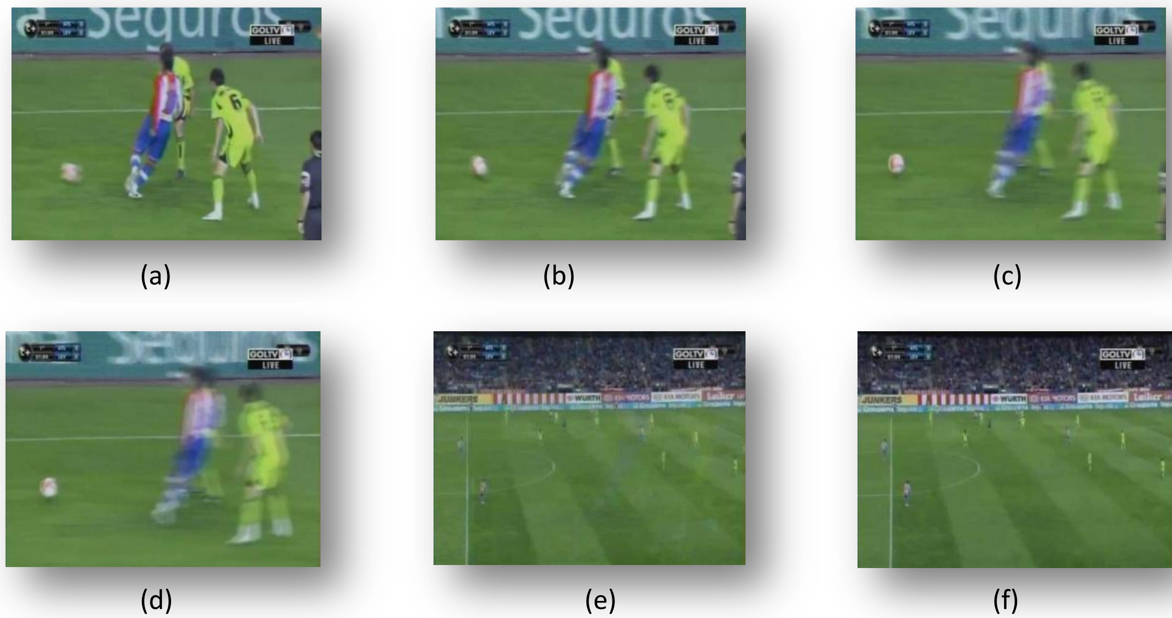
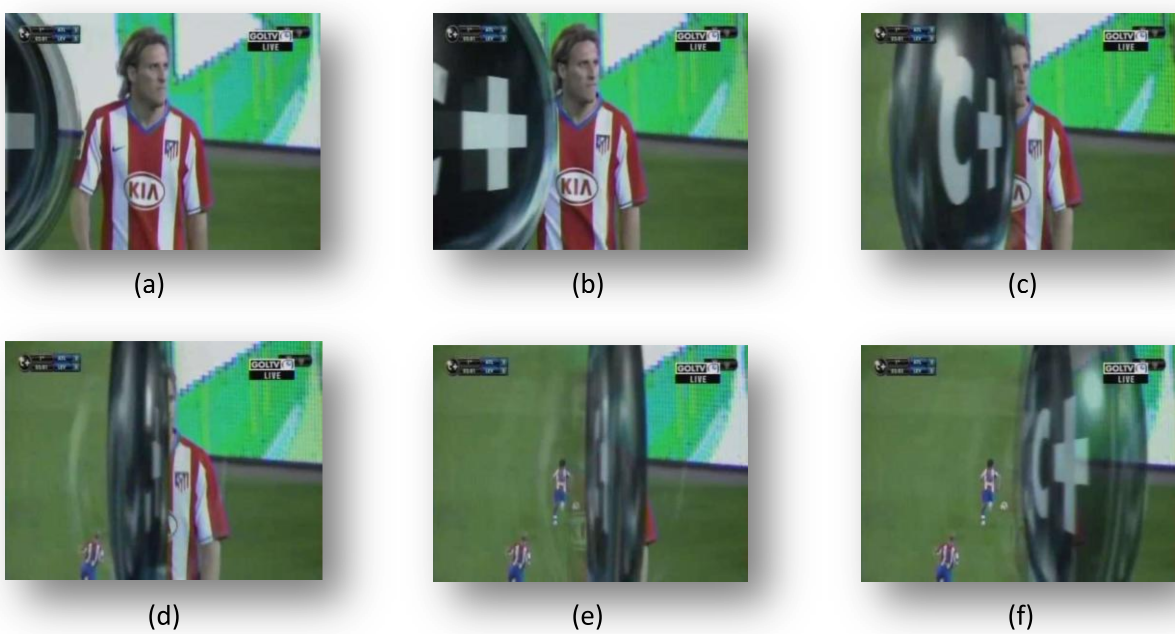


Figure (4.2): Example of instant (cut) shot transition



(g) (h) (i)

Figure (4.3): Example of gradual shot transition

4.2 Shot-boundary detection algorithm

After discussing the previous basic concepts that are very important to deal with in this chapter, we are going to explain the hierarchal steps of the shot-boundary detection algorithms simply by following the flowchart in Figure (4.4).

We use three features for shot-boundary detection in sports video as proposed in [1]:

- The difference in color histogram similarity.
- The difference between dominant colored pixel ratios of two frames.
- The dominant color pixel ratio in a frame.

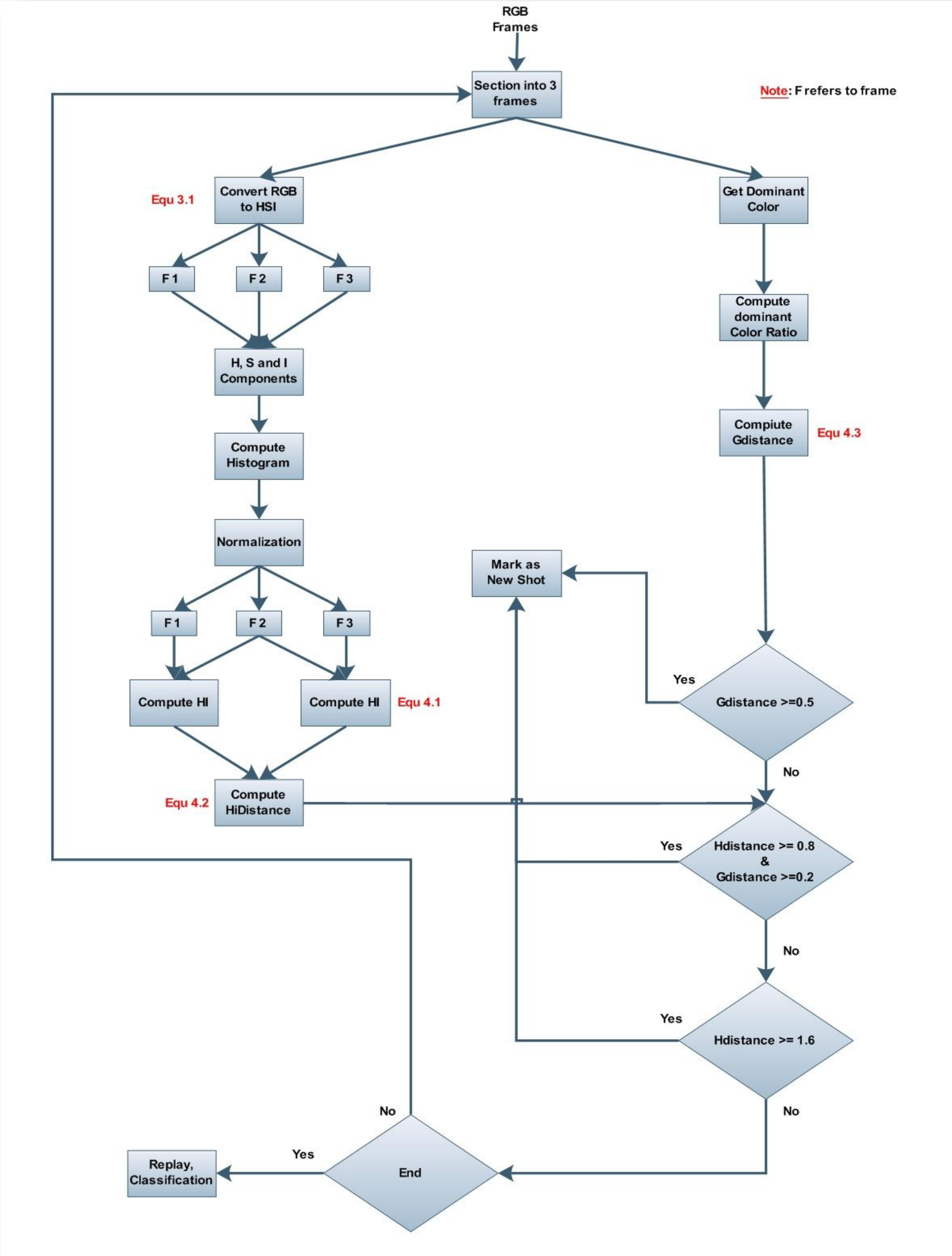


Figure (4.4): **Flowchart of the shot-boundary detection algorithm**

“Since color histograms are robust to moderate object and camera motions, we represent a frame by its color histogram, which is defined in the HSI space.
The similarity between the two histograms is measured by histogram intersection as it is found to be the most robust metric in Equation (4.1) is employed to compute the similarity between the i^{th} and the $(i-k)^{th}$ frames, HI (i, k): In the same equation, N denotes the number of color components and it is three in this case, B_m is the number of bins in the histogram of the m^{th} color component, and H_i^m is the normalized histogram of the i^{th} frame for the same color component” [1] [10].

“As explained above, the color correlation between two consecutive sports video shots will be high when field region consists of a large portion of the frames in both shots. In these cases, although there may not be a large change in global color histogram features of the two frames, there may be a recognizable difference in their field colored pixel ratios. The proposed feature, the absolute difference between dominant colored pixel ratios of two frames, is denoted by G_d and it is computed by Equation (4.3), where G_i and G_{i-k} represents the dominant color pixel ratio in the i^{th} and $(i-k)^{th}$ frames, respectively. Although the proposed feature is pixel-based, which is known to be sensitive to object motion, its integrated use with color histogram makes robustness to object motion possible” [1] [10].

Finally, we use dominant color pixel ratio in a frame. The reason to use this feature will be explained in the next section, where we discuss the design of the classifier and the adaptation of the parameters.

$$HI\ (i,k) = \frac{1}{N} \sum_{m=1}^{N} \sum_{j=0}^{B_m-1} min\ (H_i^m[j], H_{i-k}^m[j]) \quad \textbf{(4.1)}$$

$$H_d(i,k) = |HI\ (i,k) - HI\ (i-k,k)| \quad \textbf{(4.2)}$$

$$G_d(i,k) = |G_i - G_{i-k}| \quad \textbf{(4.3)}$$

4.3 Predefined Thresholds

The proposed shot-boundary detection algorithm is able to detect both instant and gradual transitions with a high detection and low false alarm rates.
The detection of gradual transitions in sports video is particularly difficult because of the high color correlation between two shots as Figure (4.3) illustrates a wipe transition in a soccer video.

"Instead of computing the difference vector for every consecutive frame pair, the comparisons are performed between frame i and frame i + placement: Unless there is a significant difference, the comparisons are only defined for frames that are placement apart. When a significant change is found, a binary search is employed to find the exact location of the difference. This method is able to find both cuts and gradual transitions. A modified version of this algorithm is obtained by comparing a range of placement values instead of one placement value" [1].

"We have observed that the optimal way to detect these types of transitions is to use a modified version of placement algorithm. In the modified placement algorithm, we compare two consecutive frames in order to detect cuts, and increment the distance between frames up to N frames to detect gradual transitions. The reflection of this method to the Equations (4.1) and (4.2) is that different k values are employed to detect cut-type boundaries and gradual. Transitions since cuts are instant transitions, k = 1 detects cuts, while we check a range of k values, k > 1 up to N = 8 to locate gradual transitions" [1].

"The detection of a shot boundary involves comparing H_d, G_d , the dominant colored pixel ratio in the i^{th} frame, with a set of thresholds. A novelty of the proposed method, in addition to the introduction of G_d as a new feature, is the adaptive change of the thresholds on H_d" [1].

In the given flowchart Figure (4.4), we use, G_d and H_d for adaptation of the thresholds because when a sports video shot corresponds long, medium, close, and logo view, they can be detected by the following thresholds values which is concluded after training.

Parameter	Value
G_d	0.5
H_d and G_d	0.8 and 0.2
H_d	1.6

Table (4.1): Parameters for shot detection learned from soccer datasets

After a learning stage, the optimum values for these thresholds can be set for each sport type. In Table (4.1), the values for these thresholds are given for soccer.

The final stage in this process is to check the shot detected videos are greater than a predefined length to prevent the occurrences of irrelevant shots.

4.4 Results

Now, we're going to present some statistics on the shot-boundary detection:

Match Name	AC Milan vs. Shakhtar	Ghana vs. Guinea	Manchester vs. AC Milan	Egypt vs. Cameron	AC Milan vs. Liverpool
Duration (mm:ss)	12:30	20:00	30:00	45:00	90:00
Correct	127	158	241	367	690
False	8	7	21	23	40
Miss	4	4	7	9	16
Recall	97 %	97.5 %	97.5 %	97.6 %	97.7 %
Precision	94 %	95.7 %	92 %	94 %	94.5 %

Table (4.2): **Shot-boundary detection results**

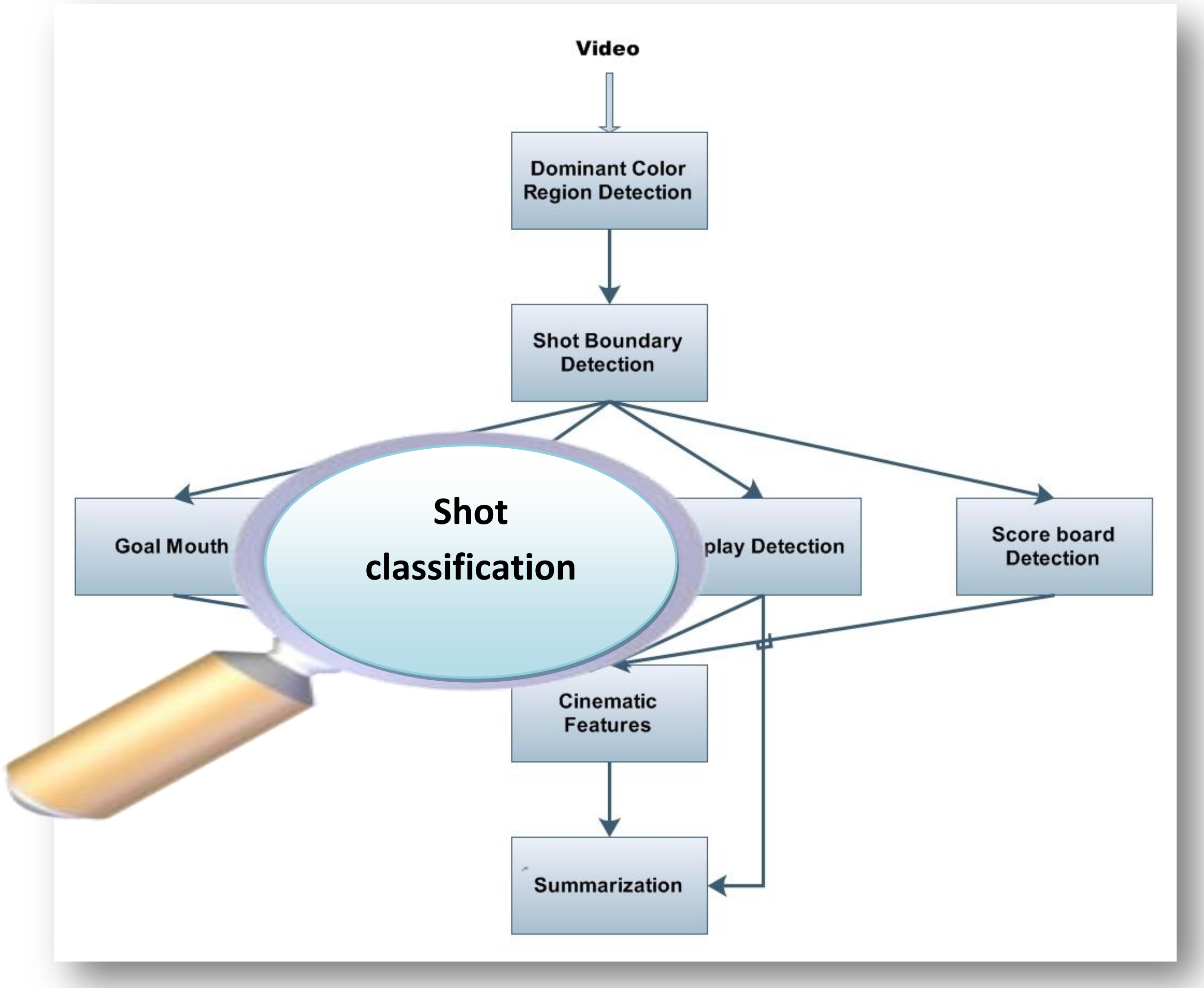
4.5 Conclusion

The proposed shot boundary detection algorithm works with high accurate result by applying the thresholds values in Table (4.1), its fast and robust, the algorithm, in its current version, is able to differentiate flashlight effects.

[5]

Shot Classification

This chapter introduces the different classes of shots: Long-shot, Medium-shot, Close-up shot and Audience (out of field) based on the camera position.



5.1 Introduction

Production crews use different shot types to convey make different scenes, which can be used for high-level video analysis in a particular domain. Cinematographers classify a shot into one of four categories long, medium, close-up and audience (out-of-field) shot classes, the definitions of which are usually domain-dependent.

In the following, we define these four classes for sports video:

Long shot: A long shot displays the global view of the field; long shots almost always display some part of the stadium, which decreases the dominant colored pixel ratio.
As shown in Figure (5.1): (a) and (b); hence, a long shot serves for an accurate localization of the events on the field.

Medium (In-field) shot: A medium shot, where a whole human body is usually visible, is a zoomed-in view of a specific part of the field as in Figure (5.1): (c) and (d).

Close-up: A close-up shot usually shows the above-waist view of a player or referee as in Figure (5.1): (e) and (f).

Audience (Out-of-field) Shot: The audience, coach, and other shots are denoted as out-of-field shots as in Figure (5.1): (g) and (h). The sequence occurrence of a close-up shots and audience (out-of-field) indicates an important event such as (goal, goal attempts …etc) during the match [1][12].

Figure (5.1): Shot types: (a, b) long, (c, d) medium (in-field), (e, f) close-up, and (g, h) out-of-field shots

The above definitions consider shot-types as functions of field region. Because field region information is available after dominant color region detection (discussed before in chapter 4) and using only the field region information that may satisfy the requirements of a particular application. We propose a shot-type classification algorithm based on dominant color region information which is used, but for more accurate classification, we should make additional step to make more robust classification.

5.2 Shot-Type Classification by Dominant Color Ratio

We use dominant color ratio in a single frame to determine the frame view-type. The proposed algorithm is based on a specific threshold (range) for grass ratio (G); which we build it offline for a particular sports type by training each class with different shots from different matches to define the specific range for the four shot type and adapt the parameters online to a particular sports type, and assigning each shot to one of the shot-types.

However, a particular sports type may not have to fit into this ideal representation as the variations in the physical specifications of the field may result in a differently shaped dominant color ratio histogram.

Classification of a shot into one of the above three classes is based on spatial features. Therefore, shot class can be determined from a single key frame or from a set of frames selected according to a certain criteria. Due to the computational simplicity of our algorithm, we find the frame grass colored pixel ratio, G, is computed, we calculate average of grass ratio for all frames, then we find the class of shot according to each class grass ratio range, but to increase the system performance we use eight key frame to compute the grass ratio.

We used an intuitive approach, where G value almost zero and most of frame is black for out-of-field view, a low G value in a frame corresponds to a close-up, while high G value indicates that the frame is a long view, and in between, a medium view is selected. Although the accuracy of the above simple algorithm is sufficient for our applications.

5.3 Shot Classification using grass ratio

We then need a set of thresholds which can distinguish the grass-ratio for the different shot types [11]. For our system we applied G_L= 0.5 to 0.75 for global (long) shot, G_M = 0.35 to 0.5 for medium shot, $G_{C/A}$ = 0.35 for close-up shot and audience shot. And using additional thresholds that can distinguish between close-up and audience shots, we use black ratio in the frame, B_A = 0.87. All of the thresholds values based on a lot of training. Using these four thresholds ratio, each frame can be classified into global, medium (in-filed), close-up or audience (out-of-filed) based on this rule in Equation (5.1).

$$\text{Frame Type} \begin{cases} \text{Global}, & G_L >= 0.5 \;\&\&\; G_L < 0.75 \\ \text{Medium}, & (G_M >= 0.35 \;\&\&\; G_M < 0.5) \;\|\; (G_M >= 0.75) \\ \text{Close-up}, & G_{C/A} < 0.35 \;\&\&\; B_A < 0.87 \\ \text{Audience}, & G_{C/A} < 0.35 \;\&\&\; B_A >= 0.87 \end{cases}$$

Equation (5.1)

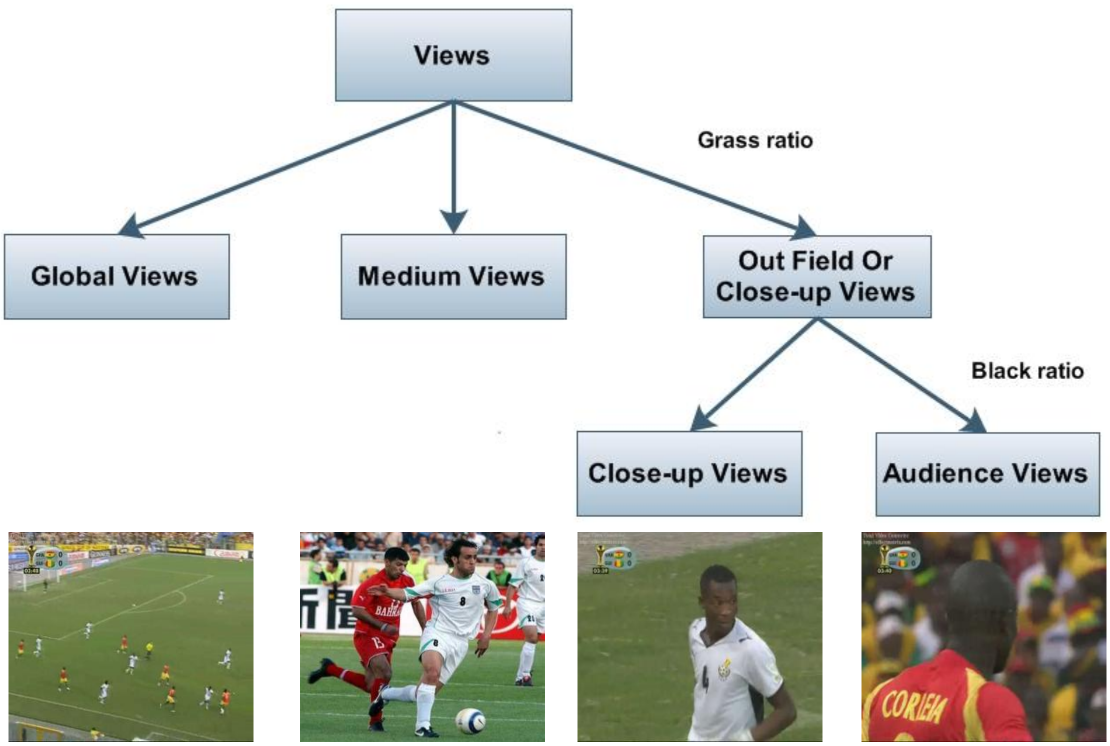


Figure (5.2): Different shot types

Grass ratio values act as the major criterion in differentiating the outfield views and infield views (long and medium views). Then the Outfield views are further categorized into close-up and audience (out-of-field) views using the black ratio value as shown above in Figure (5.2) [13]. It's relatively easier to detect the grass area. Therefore, the grass ratio value serves as the primary differentiating factor with the facilitation of roughly estimated playfield area, to improve our shot classification accuracy we should use an addition feature such as object (player, referee, ball …) size in the playfield [8]. In our shot classification we use only grass ratio and black ratio and the result of them to make a reasonable recall and precision as illustrated in the next section.

5.4. Results

Now, we're going to present some statistics on the **Shot Classification (shot type):**

Shot Type	Long	Medium	Close-up	Audience
Long	54	6	5	1
Medium	7	58	1	1
Close-up	---	3	25	2
Audience	---	1	1	3
Recall	88.6%	85.3%	78.2%	42.9%
Precision	81.9%	86.6%	83.4%	60%

Table (5.1): Confusion Matrix for Egypt vs. Cote d'Ivoire match in Africa cup 2008, it's duration 11:04 (mm:ss).

Shot Type	Long	Medium	Close-up	Audience
Long	19	3	1	---
Medium	2	27	---	---
Close-up	---	---	6	---
Audience	---	---	---	4
Recall	90.5%	90%	85.7%	100%
Precision	82.6%	93.1%	100%	100%

Table (5.2): Confusion Matrix for Tunisia vs. Cameroon match in Africa cup 2008, it's duration 6:22 (mm:ss).

Shot Type	Long	Medium	Close-up	Audience
Long	27	3	3	---
Medium	3	24	---	---
Close-up	---	---	5	---
Audience	---	---	---	13
Recall	90%	88.9%	62.5%	100%
Precision	81.2%	88.9%	100%	100%

Table (5.3): Confusion Matrix for Roma vs. Real Madrid match in Africa cup 2008, it's duration 6:51 (mm:ss).

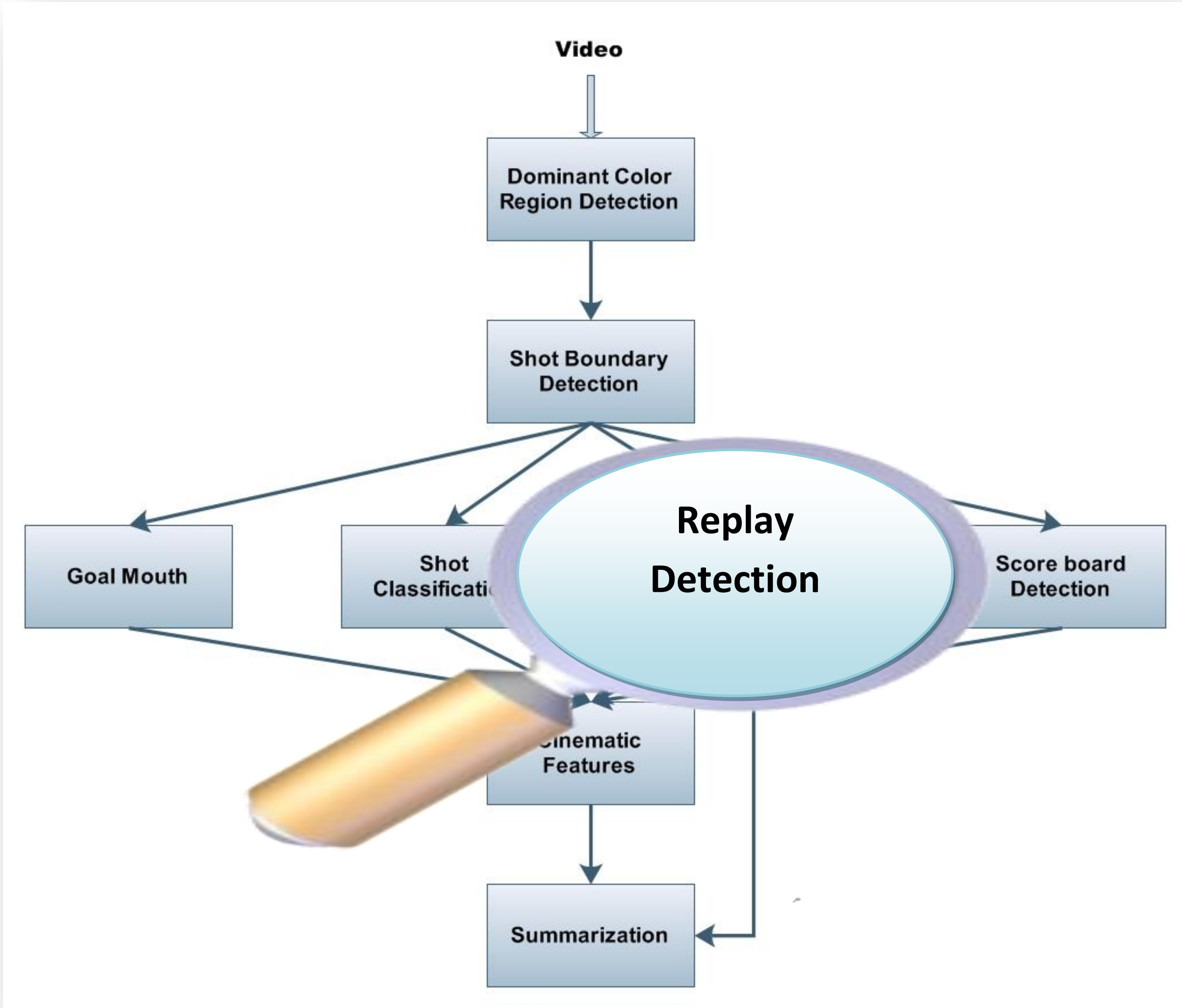
Shot Type	Long	Medium	Close-up	Audience
Long	25	3	2	---
Medium	5	23	---	---
Close-up	---	---	2	2
Audience	2	---	2	15
Recall	83.3%	82.2%	50%	79%
Precision	78.2%	88.5%	33.3%	88.3%

Table (5.4): Confusion Matrix for Netherlands vs. Italy match in Euro cup 2008, it's duration 8:10 (mm:ss).

[6]

Replay Detection

This chapter introduces the video editing techniques used to produce replay shots and the used methods to detect them.



6.1 Introduction:

In most soccer videos, exciting events are often replayed. These exciting shots normally correspond to highlights in a game, e.g., actions near the goal posts in a soccer game.

Our interest in this chapter is the detection of the replayed shots that represent exciting events in soccer videos.

6.1.1 What is meant by replay?

Replay is a video editing way, and it is often used to emphasize an important segment with a slow-motion pattern or logo appearance for once and several times.

Replay segments are often generated by two mechanisms:

1- Slow motion replay:

- Slow motion replay from the same camera.
- Slow motion replay from another camera.

Slow motion replay from the same camera can be achieved by repeating or by interpolating frames in the original sequence and playing the modified sequence at the normal frame rate. The number of repeated or interpolated frames depends on the desired speed of slow motion [14].

2- Logo based replay:

In sports video, there is often a highlighted logo that appears at the start and end of a replay segment.

6.2 Overview of the proposed system:

Replay detection is one of the important stages in soccer video summarization. The inputs to this stage are the basic units of soccer video. These units are the output shots from the shot detection stage. Once the output generated from the shot detection stage, it passed to the replay detection stage to extract the exciting events that are represented by the replay shots Figure (6.1).

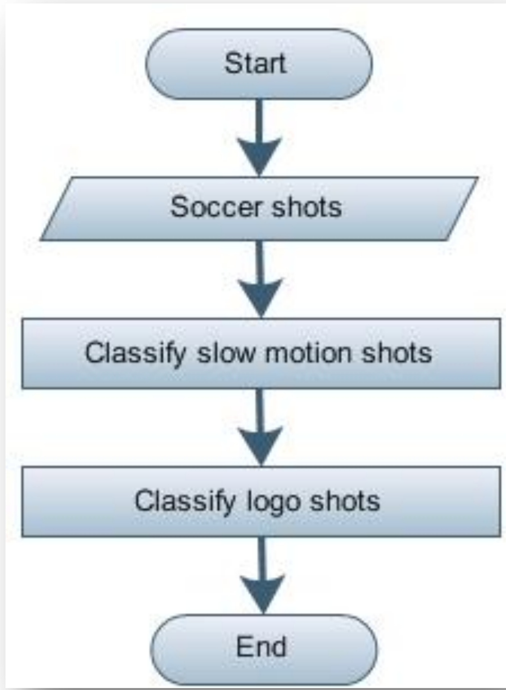


Figure (6.1): Replay Detection Flowchart.

6.2.1 Replay Classifiers:

In our proposed system we have two shot classifiers:

1- Slow Motion Classifier
2- Logo Classifier

The input shot classified into one of the below classes according to slow motion and logo classifiers. First the shot passes through the slow motion classifier to detect if it's slow motion or not. Then it passes through the logo classifier to detect if logo patterns exist in the given shot or not.

6.2.2 Generated Replay Summary:

After the inputs to the replay detection module have been processed, each input shot classified into one of the following classes:

- Slow Motion Replay Class.
- Logo Replay Class.
- Non-Replay Class.

The exciting events in soccer video exist in the first two classes, so we include all shots that classified into one of the first two classes in the summary.

6-3 Slow Motion Detection

In this section we will talk about the first part of the replay detection stage, also we will present the relevant work in slow motion detection.

6.3.1 Relevant Work

“Slow motion replay segments were detected according to human specified digital video effects. The used technique is based on the observation that many slow-motion replays are produced by repeating frames recorded from standard cameras. As a result, the image difference between pairs of adjacent frames within slow-motion replay segments frequently alters between zero and non-zero differences. Subsequently, this fluctuation is the main feature used to detect slow motion replays in these works. Obviously, this technique is not general enough to detect all types of slow-motion replays in different soccer video” [4].

6.3.2 Limitations of the current works:

There are many reasons that limit the relevant work:

1- Not all slow motion shots that produced by the different broadcasters have the same speed, so it is not general enough to recognize slow motion shots with different speeds.

2- The frames are sometimes slowed down only at the last part in a replay shot but the rest of the shot is recorded in a normal speed.

3- Some non-slow motion shots have the same fluctuation ratio as slow motion shots, so they may wrongly classified as slow motion shots.

6.3.3 Our Approach

In this section we present our approach in slow motion detection. Many difficulties faced us to detect slow motion shots: slow motion shots have different speeds from one broadcaster to another and there are non-slow motion shots that have the same characteristics of slow motion shots.

We developed our approach to be independent from the number of inserted frames (or repeated frames) that control the speed of the slow motion and can vary within a replay shot or different replay shots.

Based on the characteristics observed from slow motion shots, the camera zoomed on the action that represent the exciting event, so if any object's movement will lead to large frame variation between two consecutive frames because there are small number of the green field pixels shown in the sequence Figure (6.2), but in long and normal view shots the scene contains about 75% of the green field, so if any object's movement will result small frame variation because the movement doesn’t include large number of the green field pixels shown in the sequence Figure (6.3).

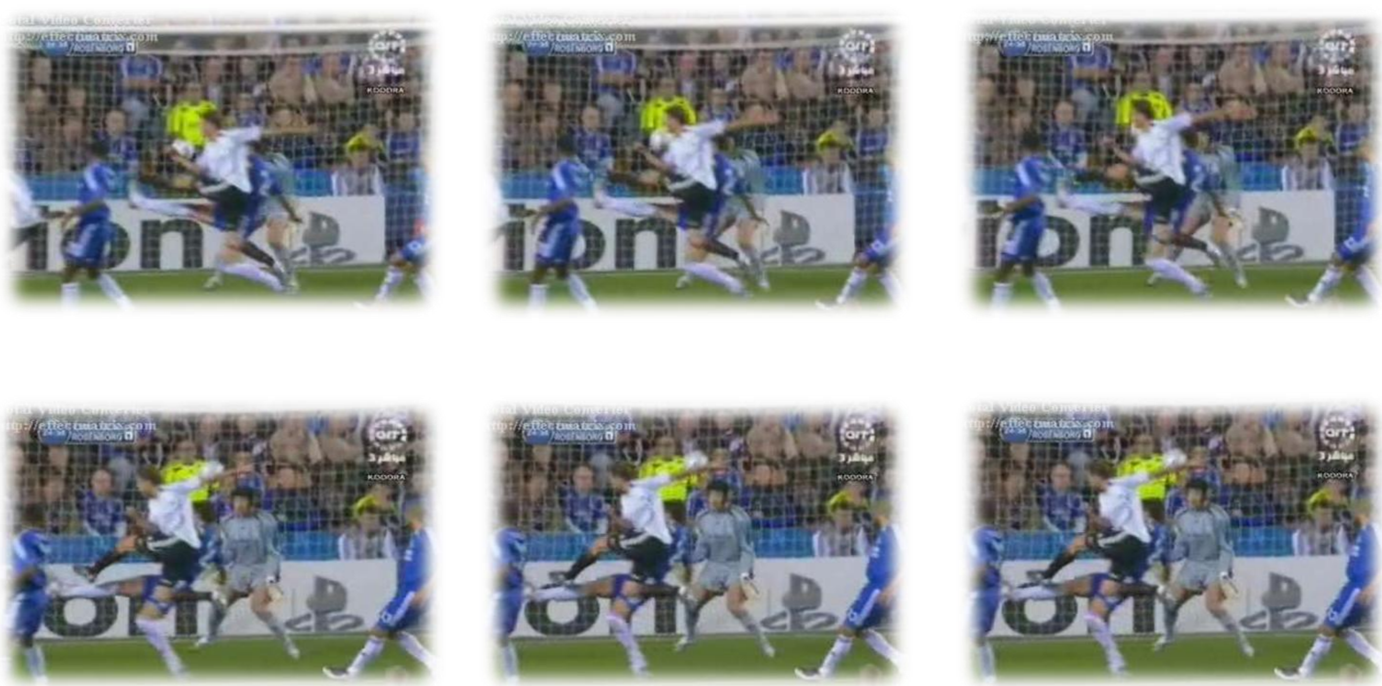


Figure (6.2): Slow motion replay sequence.

Figure (6.3): Long view sequence.

We avoid the first two problems because the approach doesn't depend on the shot speed and minimize the effect of the third one by removing short shots that have the same variation ratio of slow motion shots Figure (6.4).

6.3.3.1 How Slow Motion Detection Works?

In the slow motion detection module the input shot passes through several processes to classify it either slow motion or non-slow motion shot. Once the input in the form of soccer video shots is generated by the shot detection, it process shot by shot in the slow motion detection module.

The shot classification is done by the following four major stages Figure (6.5):

Shot Preprocessing Stage:

We observed in our training samples that the time of slow motion shots are not less than 4 seconds, from this observation we delete any shot less than 4 seconds to minimize the processing time, also this stage used to increase the system accuracy because the system may wrongly classify some close-up shots Figure (6.4) as slow motion shots.

Figure (6.4): Close-up sequence.

Usually the close up sequence time is less than 4 seconds and rarely exceeds this time. So this stage is useful to eliminate these shots before the processing.

<u>Shot Variation Measure Stage:</u>

The input to this stage is a shot where its time equals to or greater than 4 seconds. The responsibility of this stage is measuring the variation in the shot.

How we measure the shot variation?

Mainly the shot variation depends on the average frame difference between two concessive frames. We adjust the upper and lower thresholds for the variation ratio that can differentiate the slow motion shots from other shots.

<u>Shot classification stage:</u>

The input to this stage is the value measured by the shot variation measure stage. The variation measure in slow motion shots lies between 0.4 and 0.8. Therefore we can use this measure to determine if the given input shot is slow or non-slow motion shot.

6.3.4 Results:

The slow-motion detector has detected 32 of 39 replay shots, 82% recall rate, and incorrectly labeled 36 normal motion shots, 56% precision rate, as replays. The detector is sensitive to resolution and precise pixel values. The content features, such as abrupt and fast camera motions in long shots and irregular object motion in close-ups, are the main reasons for false positives.

Sequence Name	Championship	Duration (mm: ss)	# of Replay Shots	Correct	False	Miss
Milan vs. Shakhtar	Europe	10:00	6	5	5	1
Milan vs. Bayernmunich	Europe	09:00	7	5	4	2
Manchester united vs. Milan	Europe	11:00	5	5	3	0
Part-1	Africa	16:20	5	5	3	0
part-2	Africa	06:22	4	4	3	0
part-3	Africa	06:11	3	2	1	1
Part-5	Africa	06:51	3	2	0	1
Part-6	Africa	07:06	3	2	4	1
Part-7	Africa	07:18	3	2	2	1
TOTAL		79.55	39	32	25	7

Table (6.1): Slow motion detection results.

Recall = 32/39 = 82%
Precision = 32/ (32+25) = 56 %

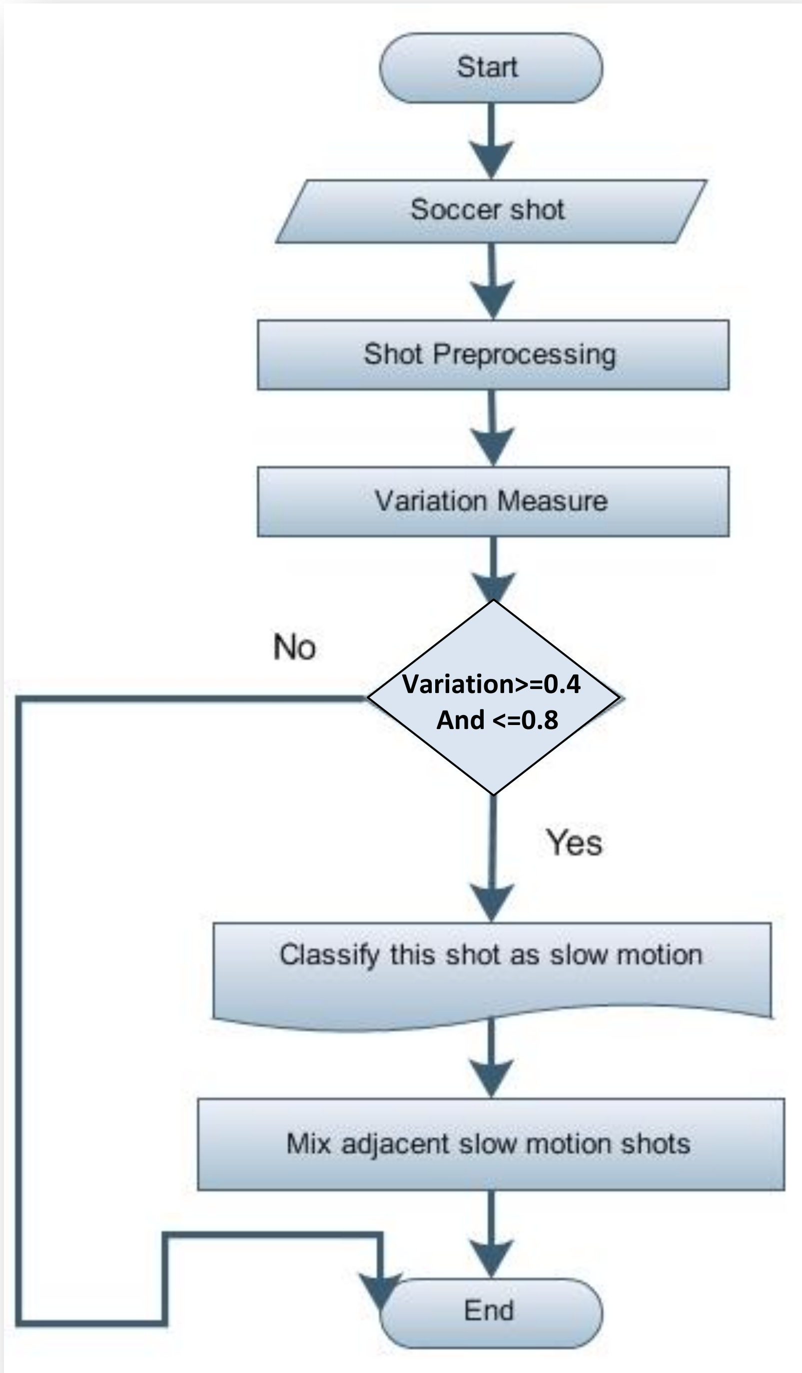


Figure (6.5): Slow Motion Replay Detection Flowchart.

6-4 Logo Detection

In this section we will talk about the second part of the replay detection stage..

6.4.1 Overview:

In sports video, there is often a highlighted logo that appears at the start and end of a replay segment that indicate an exciting event within the soccer match.

In the recent years, the broadcasters use inserted logo sequence Figures (6.6) and (6.7) as a digital video effect to replay the exciting and important events in soccer videos. In this mechanism the replay segment repeated with the original sequence speed and there is no frame repetition or interpolation [15].

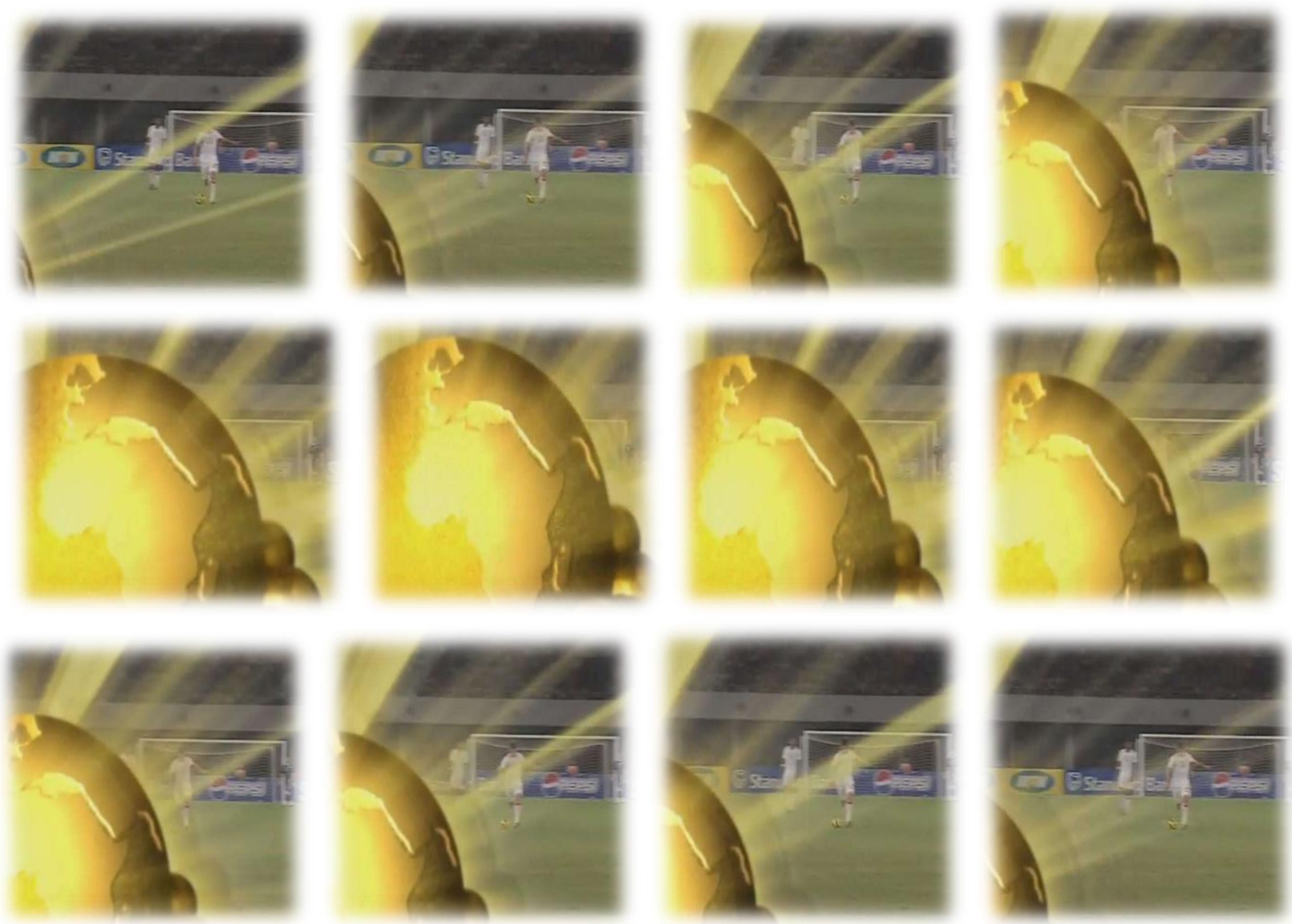


Figure (6.6): Logo sequence A.

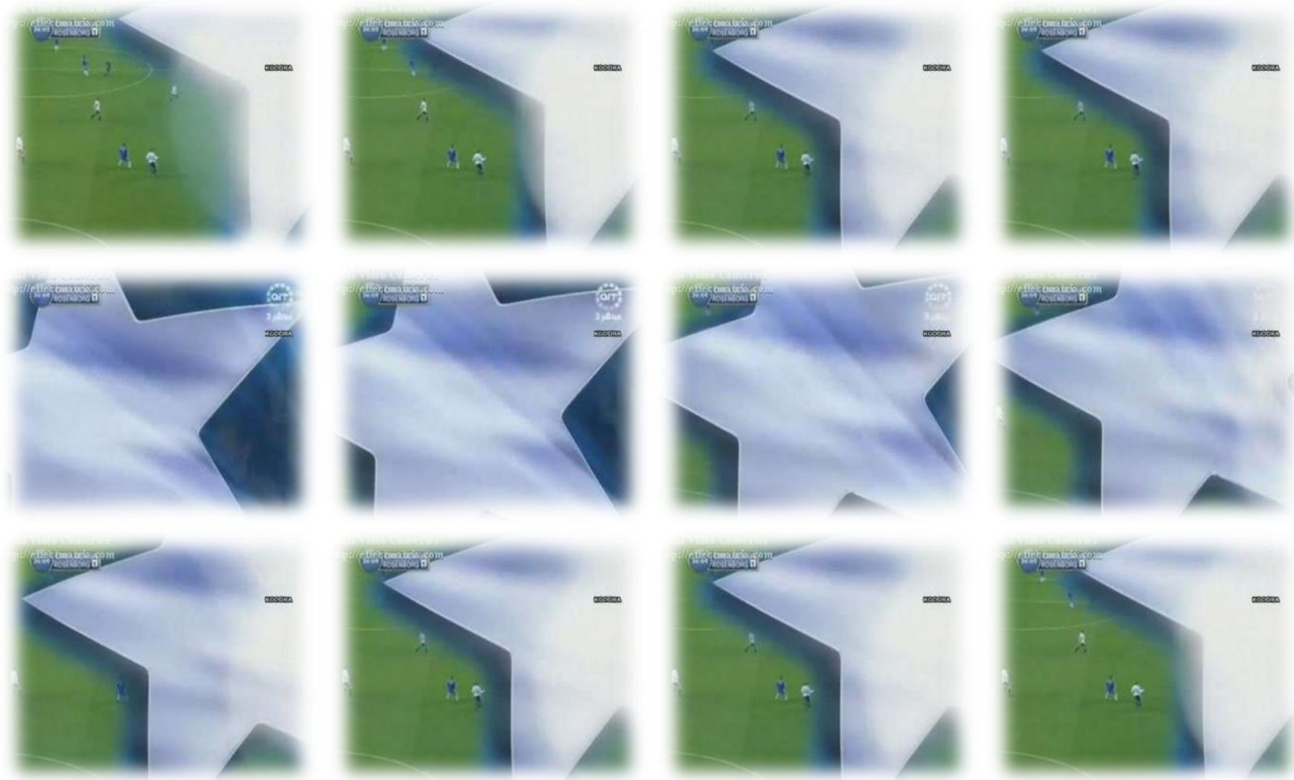


Figure (6.7): Logo sequence B.

Our aim in logo detection is the identification of a wide variety of logo used by different broadcasters.

Based on the findings, a generic logo-model can be developed as: Logo is meant to be contrasted from the background and is usually animated within 10-20 frames with a general pattern of “smallest–biggest–smallest”. Biggest contrast usually takes at least 40-50 percent of the whole frame, whereas smallest contrast is up to 10-20% [4] [12].

Unfortunately it is not easy as you expect. Some players wear a white T-shirt or one with large contrast. The system may consider these frames a logo frames by mistakes.

6.4.2 How Logo Detection Works?

In the logo detection module there are several processes passed by the input shot to determine the boundary of the replay scene that bounded by logo. There is one feature generated by the shot detection stage used in the logo detection module to enhance its performance. Logo is appeared at the beginning of the shot by the feature. This feature is used to enhance the logo detection performance as well as its accuracy because we process only a few frames at the beginning of each shot.

The shot processing is done by passing through three major stages Figure (6.8):

Calculate frame white ratio: The input shot contains a number of frames. We calculate the white ratio for each of the 35 frames in the shot to show if any of them contains a large contrast object or not. Not all frames that have a large contrast object are logo frames, so the frames are passed to the next stage.

Is it a real logo?

Usually the logo covers the entire screen Figures (6.6) and (6.7) where in non-logo frames the objects cover only a portion of the screen. We use this feature to decide if the input frame is real logo or not. Then we realized that the results need to be modified to reflect high accuracy, so we developed a new method to assure the occurrence of logo shot. This is done through taking an image of the logo for each championship or league to calculate the average color pixels histogram .Then we put a defined range to be compared with the input shot.

Add logo frame index to the array list: After we determine the logo frame we add its index to the array list. We use this array to generate the soccer video summary by producing video segments that lies between the logo frames.

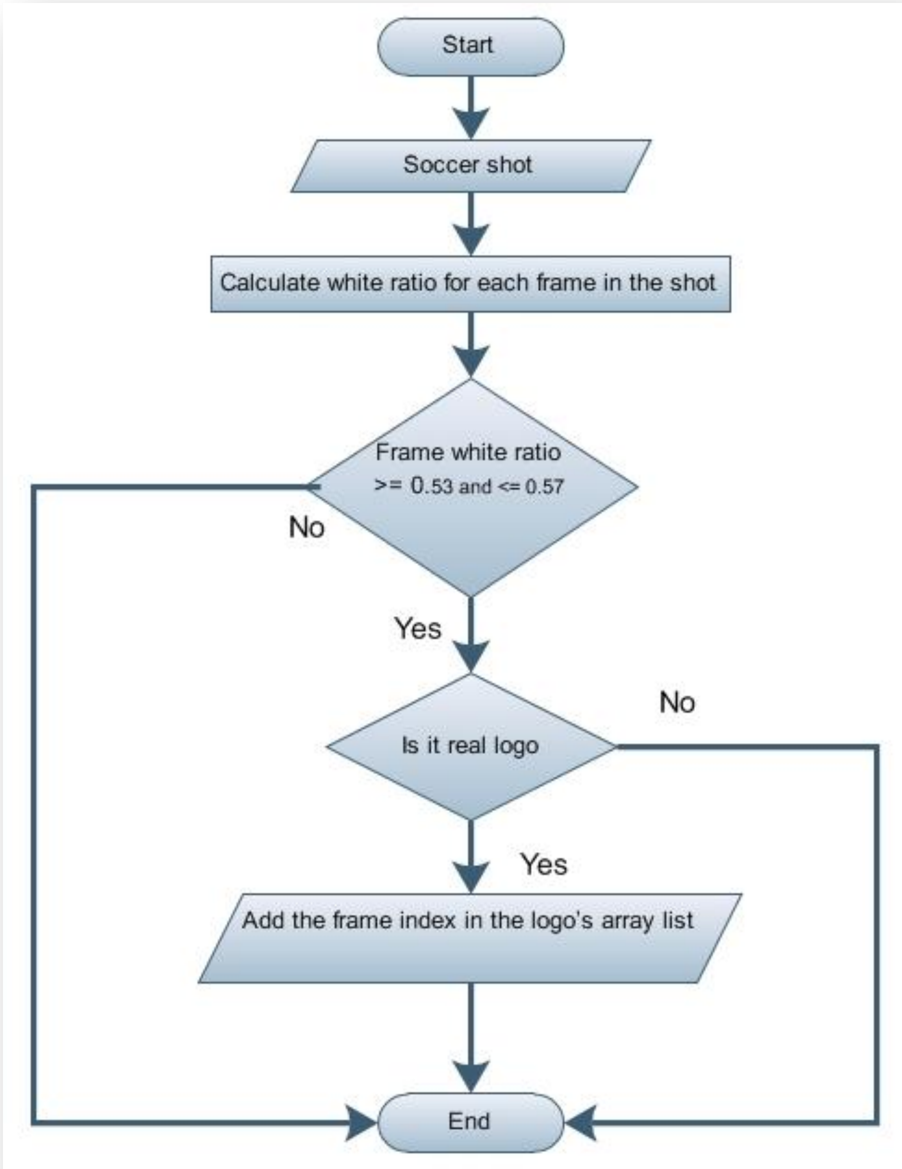


Figure (6.8): Logo Detection Flowchart.

6.4.3 Results

Now, we're going to present some statistics on the Logo Detection:

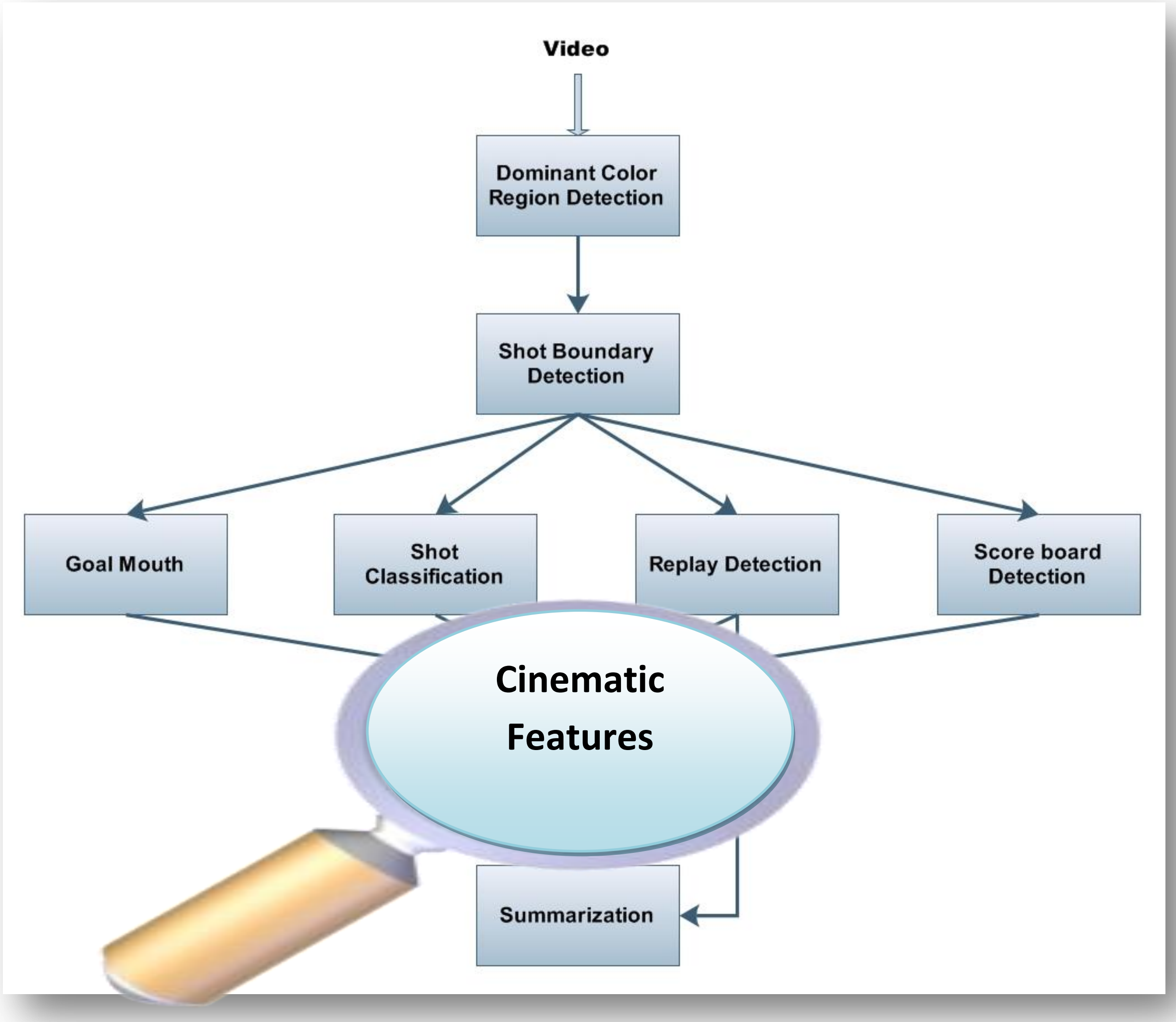
Match Name	Egypt Vs. Cote d'Ivoire	Tunisia Vs. Cameroon	Roma Vs. Real Madrid	Netherlands Vs. Italy
Duration (mm:ss)	17: 47	7:06	22:05	8:10
Correct	26	10	20	8
False	1	---	---	1
Miss	---	---	---	---
Recall	100%	100%	100%	100%
Precision	96.3%	100%	100%	88.9%

Table (6.1): Evaluation of Logo based replay.

[7]

Cinematic Features

This chapter introduces the definition of cinematic features and how to extract them to help in the detection of the important events.



7.1 Introduction

"Cinematography defines the rules, techniques, and conventions in filmmaking. It specifies the most general and basic filming techniques, such as the rules for positioning the main objects on the screen and for setting the camera locations, as well as some tailored representations for specific domains, e.g., the use of slow-motions in sports. The video features that result from these generic and domain-specific rules and conventions are referred to as cinematic features, and their efficient extraction and analysis makes fast and semantic video processing possible. one of the main disadvantages of all the cinematic features are that they aren't generic in the sense that they are specific for only one sport and dependent of the cinematic style employed by a particular broadcasting crew" [1][18].

The cinematic features are considered a sequence of shot-types (shot classification), slow-motion replays, logo based replay, score board and goal mouth detection. As shown in Figure (7.1), these features allow us to detect goals, goal attempts, attacks and others events (fouls, injury and offside).

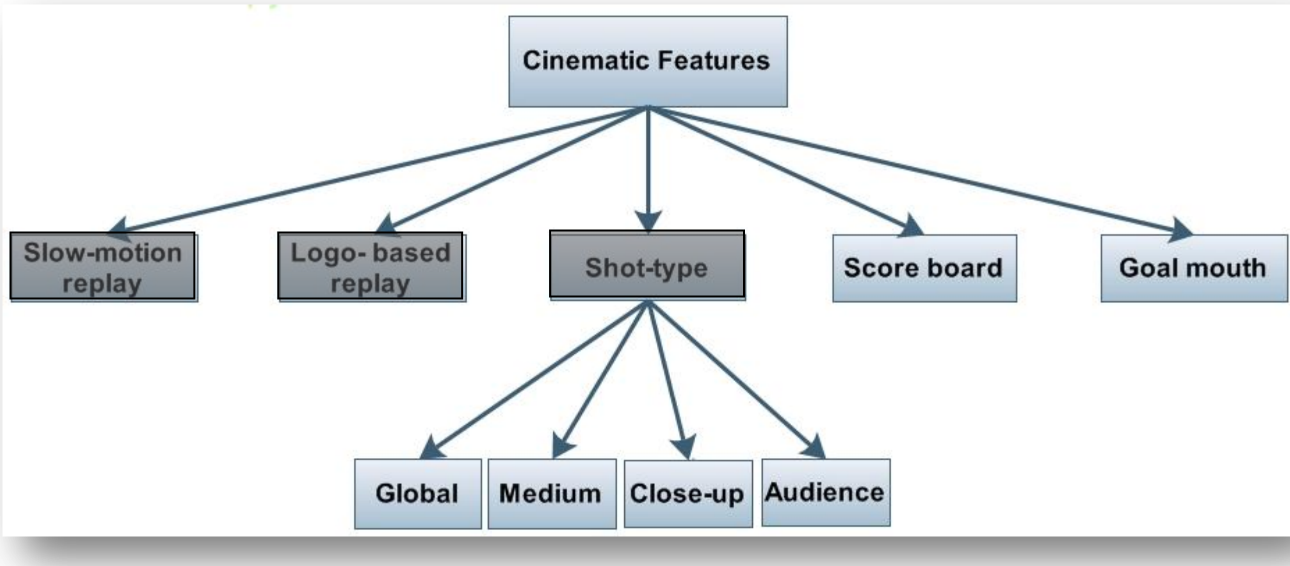


Figure (7.1): The Cinematic Features.

This chapter expounds different sections. Section 7.2 will cover score board detection. Section 7.3 will cover goal mouth detection. Section 7.4 soccer summarization. Section 7.5 results for event detection.

7.2 Score board Detection

"The score board is a caption region distinguished from the surrounding region, which provides the information about the score of the game or the status of the players. Here, we combine the dynamic and static properties to detect the caption region" [6].
We make use of the fact that the caption often appears at the bottom part of image frame for a short while and then disappears almost appear for 5 second. So the abrupt intensity change at the bottom part of the adjacent frames is used to detect the appearance of the caption.

Figure (7.2): The score board caption region indicates different types of the existence of the board.

The appearance of caption is quite common in the soccer video and usually accompanies the occurrence of shooting event. If the caption is detected with enough confidence, it can undoubtedly provide the inference of goal event, because after every scored goal the score board is displayed. One challenge of the score board detection is the different styles of caption region of the score board as shown in Figure (7.2).

Our method convert each frame to binary image (black and white) and then calculate the lower half black and white ratio after that check if this frame contains score board based on black and white ratio this is unique for every championship or league and use the duration of the score board appearance (around five sec = 125 frame) to increase the accuracy of score board detection, there are other algorithms [16].

For lack of the usage of domain prior knowledge, most of the detection methods cannot achieve the stable and robust performance, so in our system performs the score board detection based on the following domain knowledge about the score board of soccer video. Firstly, the score board usually appears in the lower half of frame image and keeps a certain distance to the image boundary. Secondly the score board region often has brighter gray levels representation than other regions.

7.2.1 Results

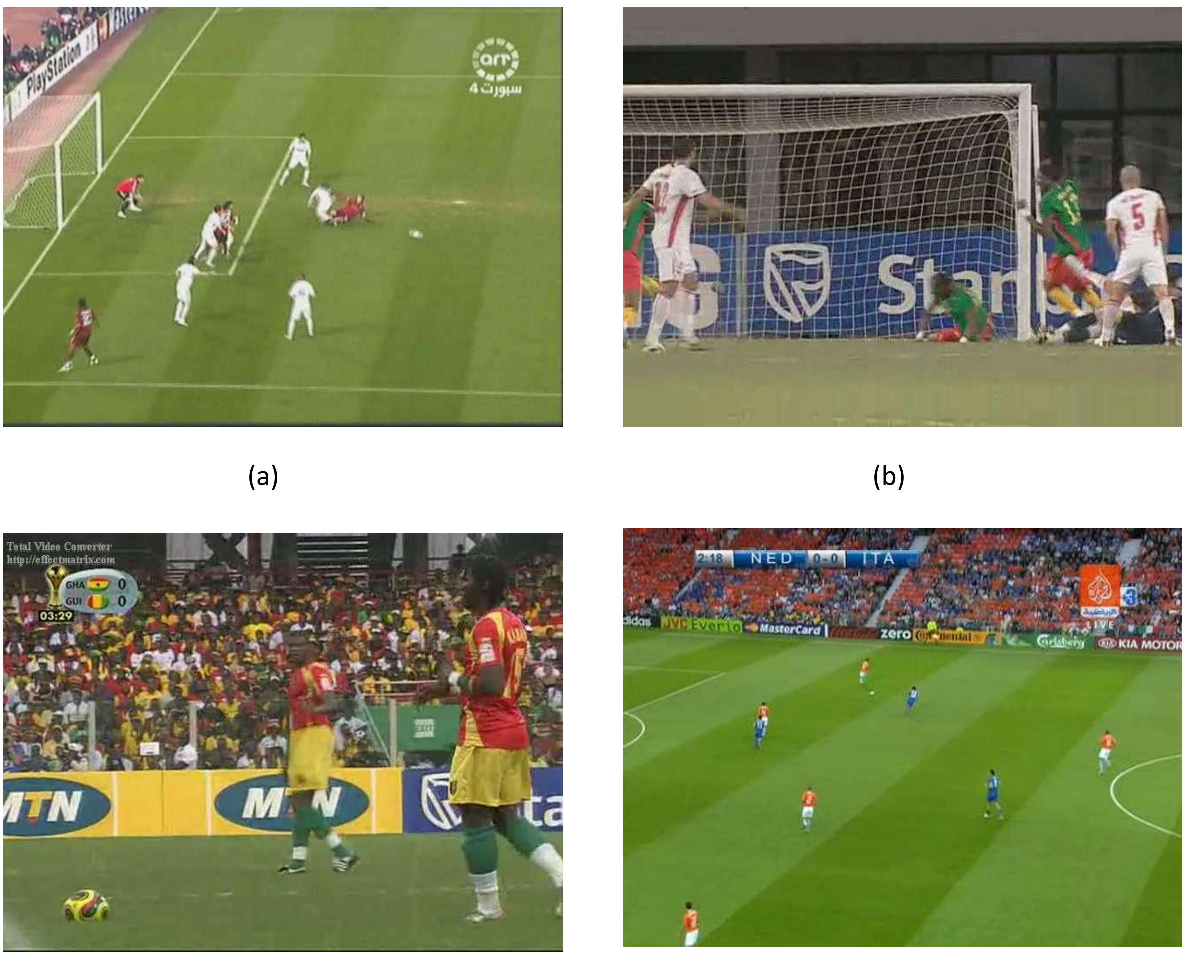
Now, we're going to present some statistics on the Score board Detection:

Match Name	Egypt Vs. Cote d'Ivoire	Ghana Vs. Guinea	Tunisia Vs. Cameroon	Roma Vs. Real Madrid
Duration (mm:ss)	17: 47	2:41	6:17	6 :53
Correct	2	1	1	2
False	1	---	---	1
Miss	---	---	---	---
Recall	100%	100%	100%	100%
Precision	66.7%	100%	100%	66.7%

Table (7.1): Evaluation of Score board from different matches.

7.3 Goal-Mouth Detection

For soccer video, the goal-mouth scenarios can be selected as the highlighted candidates, for the reason that most of the exciting events occurs in the goal mouth area such as the goal, shooting, penalty, direct free kick, etc as shown in Figure (7.3): (a) and (b). On the other hand, the non-goal-mouth scenarios often consist of the dull passes in the midfield, defense and offense or some other shots to the audiences or coaches, etc, which are not considered as exciting as the former, in Figure (7.3) (c) and (d).
So we managed to extract goal-mouth scenarios from the soccer video as the highlighted candidates. Figure (7.3) shows two scenarios which illustrate the differences between a goal-mouth shots and a non-goal-mouth shots [17].



(a) (b)

(c) (d)

Figure (7.3): A goal-mouth appearance indicates potential exciting play in (a) and (b), but in (c) and (d) are non-goal-mouth that illustrate non-exciting play.

To detect goal-mouth scenarios, we first find the goal post and its crossbar by search for white color, then search for an intersection point between the goal post and the crossbar. And to increase the accuracy of this algorithm of goal-mouth detection we use white ratio threshold and count of goal-mouth threshold to minimize the error as in Figure (7.4).

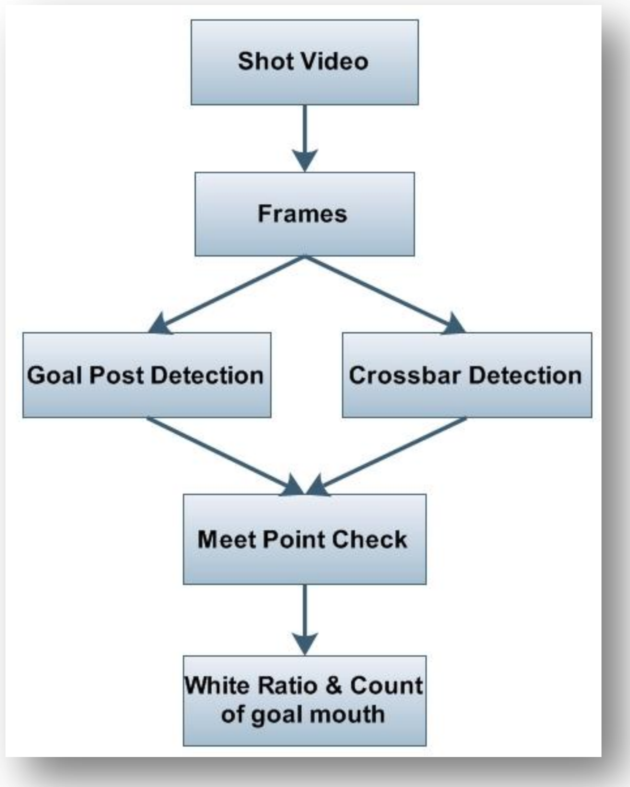


Figure (7.4): Goal Mouth Detection Flowchart .

7.3.1 Vertical Goal Post Detection

The two vertical goal posts are distinctively characterized by their vertical strips of white and grow connected pixel gray values of white. For each frame, firstly search for a white pixel between a gray value of Whitemin =180, and Whitemax =255. Because we are primarily looking for a vertical strip of white. As shown in Figure (7.5), Then take 80 pixel under this pixel (change only height, in the same column but in different rows). Check if 70 of 80 pixels are white, it will be goal post detection. We observed that on cases when the conditions failed to detect the two vertical goal post such as (the play is at a corner area), so we detect one vertical goal post as shown Figure (7.6).

Figure (7.5): Different View of Vertical Goal Post .

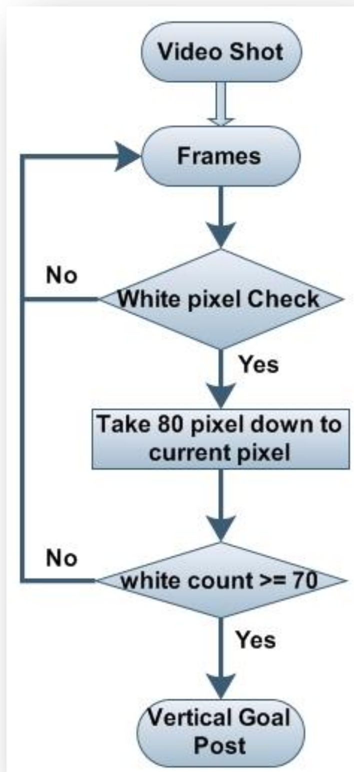


Figure (7.6): Vertical Goal Post Detection Flowchart.

7.3.2 Horizontal Cross-bar Detection

"In applications such as event detection, the detection of the two vertical goal posts isn't sufficient as a facilitating hint for possible exciting play. We need the extra step to accurately isolate the position of the cross-bar. Expectedly, overall recall performance for goal-mouth appearances will drop as a result of the additional processing and conditions therein" [7].

For horizontal cross-bar detection is connecting the two topmost endpoints of the vertical goal post. A narrower range of grays of white is defined as it is defined for the vertical goal post, for each frame, firstly search for a white pixel between a gray value of Whitemin =180, and Whitemax =255. Because we are primarily looking for a horizontal strip of white. As shown in Figure (7.7), Then take 80 pixel next to this pixel (change only width, in the same row but in different column). Check if 70 of 80 pixels are white, it will be cross-bar detection as shown in Figure (7.8).

Figure (7.7): Different View of Horizontal Cross-bar.

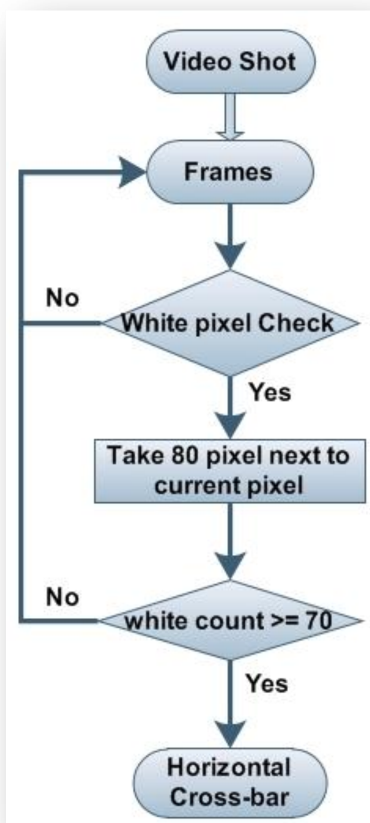


Figure (7.8): Horizontal Cross-bar Detection Flowchart.

7.3.3 Results

Now, we're going to present some statistics on the Goal mouth Detection:

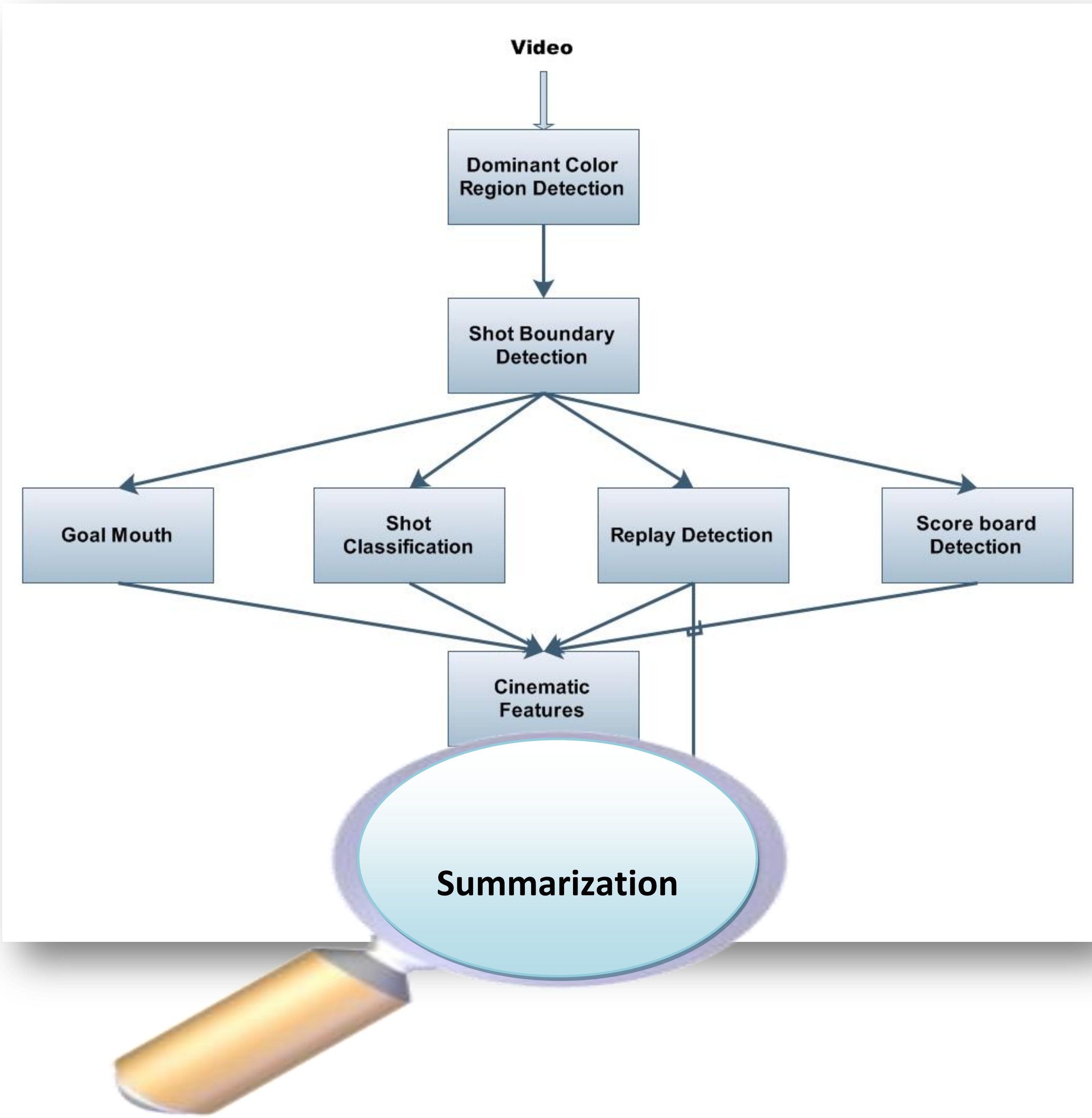
Match Name	RomaVs. Real Madrid	TunisiaVs. Cameroon
Duration (mm:ss)	6:53	6:17
Correct	10	7
False	10	12
Miss	1	---
Recall	91%	100%
Precision	52.4%	36.9%

Table (7.2): Evaluation of Goal mouth detection from different matches.

[8]

Event Detection

This chapter introduces the final stage of our proposed system using all the previous extracted information to produce a summarized video segment.



8.1 Introduction

From Chapter 4 to 7, we presented the visual feature detection (cinematic features) algorithms for soccer video analysis. Now this chapter will use all of those cinematic features to detect soccer events: Goal, attack and other events (fouls, injury and offside). As we said before cinematic features will be used as a sequence, so we use the logo-based replay to be the start of features sequence. Because of the robust detection of logo-based replay with high recall and precision over slow-motion replay then the other cinematic features will be defined as it will be discussed later in this chapter.

The users of our system can choose which type of events they want to see in the content of summarization video by selecting from different event detection types (goal, attack and other events).

8.2 Soccer Summarization

When we are talking about soccer game, we can refer to a continuous sports which mean that if there is an existence of such a break during the match, it can an indicator of the occurrence of important event such as (goal, goal attempts shots, red/yellow card, fouls and offside). Therefore the summarization process which we aimed for can be recognized by a combination of these events, for example the summarized segment may contains only the goal shots, goal attempts or penalty shots that can be described as important events [1].

In this approach, we are going to highlight the most important events such as (Goals, Goal attempts), facilitating the process of automatic match, save the viewer’s time, and introduce the technology of computer-based summarization into sports field.

8.2.1 Goal Event Detection

A goal is scored when the whole of the ball passes over the goal line between the goal posts and under the crossbar; but, it is difficult to verify these conditions automatically and reliably by the state-of-the-art video processing algorithms. Instead, we exploit our observation about the broadcasters' consistent use of a certain pattern of cinematic features after the goal events. The occurrence of a goal event leads to a break in the game. During this break, the producers convey the emotions on the field to the TV audience and show one or more replay(s) for a better visual experience. The emotions of players are captured by one or more close-up views of the actors of the goal event, such as the scorer and the goalie, and by shots of the audience celebrating the goal. Furthermore, several slow-motion replays of the goal event from different camera positions are shown. Finally, the restart of the game is usually captured by a long shot. Figure (8.1) illustrates the instantiation of such a pattern of sequence of images for the first goal of Netherlands in Netherlands vs. Italy soccer match in euro 2008 [1].


25:21 NED 0-0 ITA
LIVE

(a)


25:31
LIVE

(b)


25:33
LIVE

(c)


26:00
LIVE

(d)


26:04
LIVE

(e)


26:12
LIVE

(f)

(g)

Figure (8.1): An example goal broadcast (the temporal order is from (a) to (g)): (a) long view of the actual goal play, (b) player close-up, (c) audience, (d) the first replay, (e) the second replay, (f) the third replay, and (g) long view of the start of the new play.

8.2.1.1 Goal event features

We define the following features for the cinematic goal template that occurs between the long shot resulting in the goal event and the long shot that shows the restart of the game:

- **Duration of the event shots**: duration between two replay logos, its important feature for classifying events to goal, goal attempts, attacks and other events such as (fouls injury… etc). Due to goal events the duration no less than 16 and no more than 40 seconds.

- **The occurrence of at least one close-up shot**: This is a shot of a close-up of a player who scored the goal.

- **The occurrence of at least one audience / out-of-field shot**: This shot may either be out-of-field view of the audience or view of coach.

- **The existence of at least one slow-motion replay shot**: The goal play is always replayed one or more times from different camera positions.

- **The existence of at least one goal mouth shot**: The goal mouth is always detected one or more times during the goal event.

- **The occurrence of at least one global shot:** This shot a global view to define the restart of the match after a replay (break).

- **The existence of score board (texture) shot**: The score board is displayed after the repeating of the scored goal.

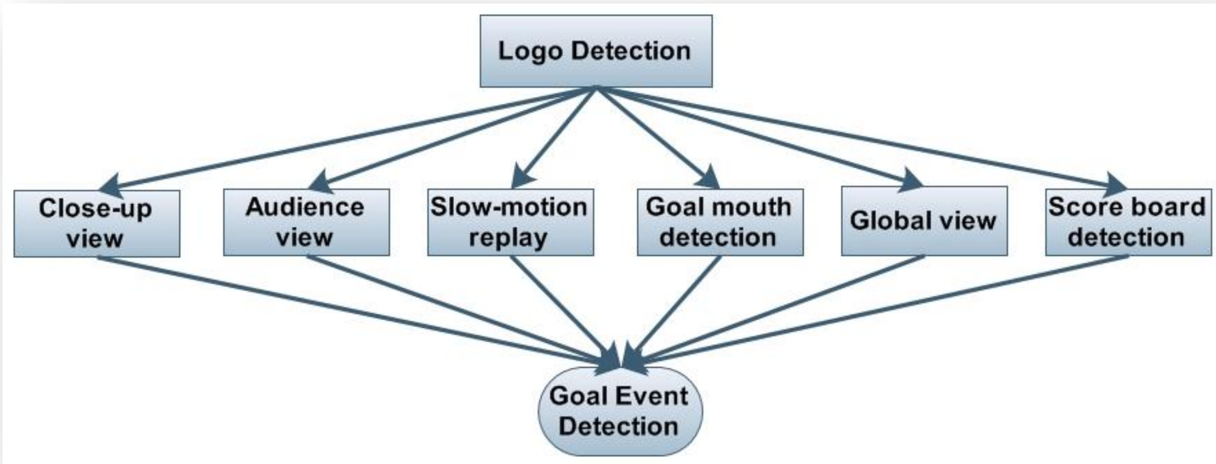


Figure (8.2): Soccer goal event detection.

In order to locate the template positions of goal event detection, we use logo-based replay shots. For every detected pair of replay logos (one for start replay shot and the other for end replay shot) to highlight an important events, after that the system finds close-up shots, audience shots, slow-motion shots to replay a goal from different camera positions, goal mouth appearance during the actual goal play and replays, global view to define the restart of the match after replay and finally the system find the score board to illustrate the result to TV viewers after scored goal. The features are defined above in Figure (8.2), are verified to detect goals.

8.2.2 Attack Event Detection

Attack events may also match a lot of goal event features, although not as consistently as goals. The addition of attack events in the summaries may even be desirable since each of these events consists of interesting shots. Therefore, the recall rate for this algorithm is much more important than the precision rate, since the users will not be tolerant to missing goals, but may enjoy watching interesting non-goal events. In the following, we describe the algorithm for the detection of attack events [1].

8.2.2.1 Attack event features

- **Duration of the event shots**: duration between two replay logos, due to attack events its duration no less than 10 and no more than 30 seconds.

- **The occurrence of at least one close-up shot**: This is a shot of a close-up of a player who made the attack event.

- **The existence of at least one slow-motion replay shot**: The attack event is always replayed one or more times from different camera positions.

- **The existence of at least one goal mouth shot**: The goal mouth is always detected one or more times during the attack event.

- **The occurrence of at least one global shot:** This is a shot of a global view to define the restart of the match after a replay.

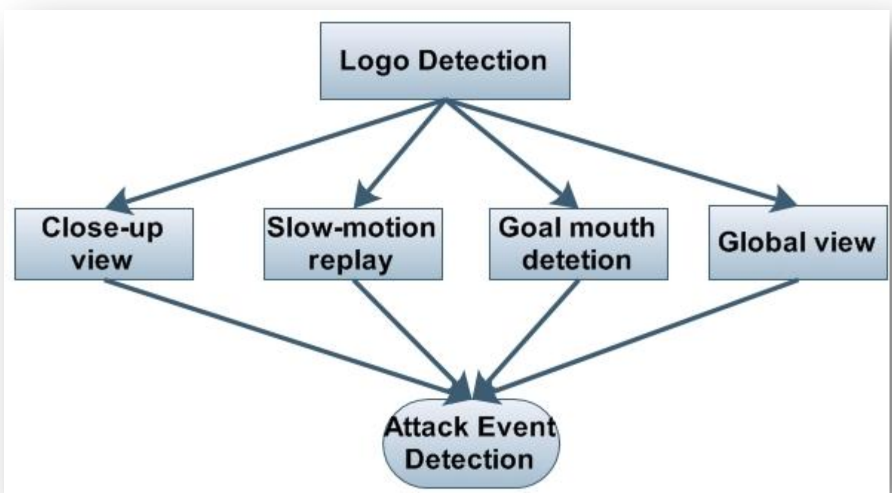


Figure (8.3): Soccer attack event detection.

In order to locate the template positions of attack event detection, we use logo-based replay shots. For every detected pair of replay logos to highlight an important events, after that the system finds the close-up shots define the player who made the attack event, slow-motion shots to replay the attack event from different camera positions, goal mouth appearance during the actual attack event and replays and finally the system find global view to define the restart of the match after a replay. The features are defined above in Figure (8.3), are verified to detect attack and goal attempts.

8.2.3 Other Event Detection

Other interesting events such as (fouls, cards, injure and offside). The addition of these events in the summaries may even be desirable since each event consists of interesting shots. Therefore, more of users may enjoy watching interesting fouls and offside events.

8.2.3.1 Other event features

- **Duration of the event shots**: duration between two logos, due to other events its duration no less than 4 and no more than 15 seconds.

- **The occurrence of at least one close-up shot**: This is a close-up shot of a player who made the event (fouls, injury and offside).

- **The existence of at least one slow-motion replay shot**: The fouls, injury and offside play are always replayed one or more times from different camera positions.

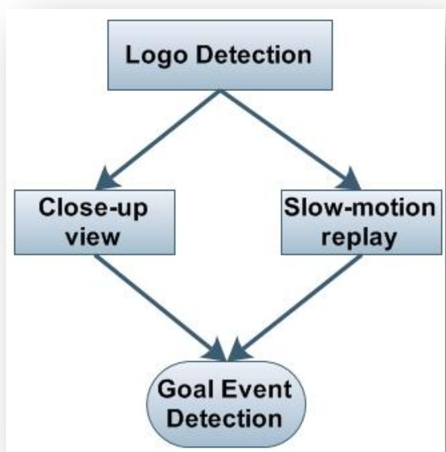


Figure (8.4): Other soccer event detection.

In order to locate the template positions of other events detection, we use logo-based replay shots. For every detected pair of logos to highlight an important events, after that the system finds the close-up shots that define the player who made an interesting events and slow-motion shots to replay an event from different camera positions. The features are defined above in Figure (8.4), are verified to detect other events (fouls, cards and offside).

8.3 Results

Now, we're going to present some statistics on the **Soccer Summarization (event detection):**

Event Detection	Goal	Attack	Other event
Goal	3	---	---
Attack	---	4	---
Other Events	---	1	5
Recall	100%	100%	83.3%
Precision	100%	80%	100%

Table (8.1): Confusion Matrix for Egypt vs. Cote d'Ivoire match in Africa cup 2008, its duration 17: 47 (mm:ss). In this video segment Egypt scores 2 goals and Cote d'Ivoire score 1 goal.

Event Detection	Goal	Attack	Other event
Goal	1	---	---
Attack	---	2	---
Other Events	---	---	1
Recall	100%	100%	100%
Precision	100%	100%	100%

Table (8.2): Confusion Matrix for Tunisia vs. Cameroon match in Africa cup 2008, its duration 7:14 (mm:ss). In this video segment Tunisia scores 1 goal.

Event Detection	Goal	Attack	Other event
Goal	1	---	---
Attack	---	2	1
Other Events	---	1	4
Recall	100%	66.7%	80%
Precision	100%	66.7%	80%

Table (8.3): Confusion Matrix for Roma-Real Madrid match in championship league 2008, its duration 22:05(mm:ss). In this video segment Roma scores 1 goal.

Event Detection	Goal	Attack	Other event
Goal	2	---	---
Attack	---	---	---
Other Events	---	---	2
Recall	100%	100%	100%
Precision	100%	100%	100%

Table (8.4): Confusion Matrix for Netherlands vs. Italy match in Euro cup 2008, its duration 8: 10 (mm:ss). In this video segment Netherlands score 2 goals.

[9]

Conclusion and Future works

This chapter summarizes our works and ideas which are introduced previously and our plans to increase the base of our proposed system.

9.1 Conclusion

We presented the main purpose of our application including our motivations, scope of work and our proposed system. Then we provided the reader with necessary background information that is needed for subsequent discussions. After that we introduced some technologies applied, although briefly, provides basic ideas in this field. The study of these techniques is fundamental to understand the next chapters.

And we presented the applied methods of dominant color region detection, numerous examples are provided to keep a clear focus on the utility and the promises of these techniques. Then we developed a new algorithm to detect the dominant color. Afterwards, we discussed the most likely applied techniques in shot-boundary detection. Also we found that some parameters must change according to the given datasets and this offers an intuitive, solid base for experimentation. And our next step is to categorize the produced video shots which came from shot-boundary detection into four classes: Long-shot, Medium-shot, Close-up shot and Audience (out of field) shot according to their camera positions. We presented the importance of replay detection process including slow motion detection and logo detection in terms of important events. Replay segments are good indicators for exciting events in any sport video.

We introduced a comprehensive foundation for event detection technique. Then solid concepts and techniques introduced a powerful set of tools for extracting features of interest. After that we highlighted the concept of the summarization process as a collection of features which are extracted from the cinematic feature process. Afterwards, we classified the important event according to the sequence of cinematic features in terms of goal event, attack event and other event.

The reader of this project will reach a level of understanding which is the foundation of most work currently underway in this field. This area of soccer video processing is a dynamic field and new techniques and applications are reported routinely in professional literature and in new product announcements. For this reason, the topics included in this chapter served the purpose of introducing the state of art in soccer video summarization techniques. In addition to enhancement, this project served the purpose of introducing a number of concepts such as dominant color detection, shot boundary detection, shot classification, replay detection and goal detection that can be used for other applications.

9.2 Challenges

Now, we introduce the solved challenges that we faced while implementing our proposed system in order to inform any beginners in this field to ease discover and prevent any problem occurred.

- **Dominant Color Detection:**

There are numerous variations which affect the degree of dominant color detection. We can mention the effect of changing the type of sport which led to change some thresholds according to the selected sport to truly detection. Also the stadium condition such as (lighting and shadows) can be considered as another reason which makes some troubles in the color region detection and we must be aware of the environmental factors such as (sunset, shadows and snow) and the field properties within the stadium which have worse effects than the mentioned factors [1].

- **Shot Detection:**

We must introduce the most significant factors that affect the accuracy of shot-boundary detection because it is a critical stage in the whole system which all other stages follow it. Shot transition is an example of these factors especially when it corrupted by the occurrence of camera or object motion which make confusion in this process. The object motion can easily be fast because we are talking about a soccer game which means that all team players must move rapidly during the match that makes some troubles and cause wrong detection of shot boundary. In addition the movements of camera zooms present another factor that effect this process, and this is an essential operation that must be done to show different views of the game.

- **Shot Classification:**

The only challenge in this stage is how to separate close-up shots and audience (out-of-field) shots, because those two types of view are joined together into one view type in most papers.

- **Slow motion detection:**

“A challenge is that the magnitude of motion varies among different types of shots. For example, in soccer video, global view shots usually have relatively much smaller motions than close-up shots, which are usually used to track motions of individual key players. It is possible that the motion of a normal global view shot may be smaller than a slow-motion replay of a close-up shot, and the motion of normal close-up shot may be equals to the slow-motion replay of a medium shot” [5].

“In addition the slow motion detector is sensitive to resolution. The content features, such as abrupt and fast camera motions in long shots and irregular object motion in close-ups, are the main reasons for false positives” [1].

- **Cinematic features:**
  - Detection of score board is different from championship to championship or from league to league.
  - In the detection of goal mouth, we realized that, before finding the goal mouth boundary, we must look for their components which are vertical goal post and horizontal crossbar by searching for the white color. After that we are looking for the intersection point between vertical goal post and horizontal crossbar.

- **Soccer Summarization:**

  Classification and highlighting the most important events such as Goals, attacks and other events (fouls, injury, offside). We enable the user of the proposed system to choose what events (goal, attack and other event) are to be generated in the summarization video.

- **Video file:**

Video is a sequence of images; the resolution of the image plays a significant role in error occurrence. Low resolution quality below 640*480 will lead to more error, because there will be noise in the image, this noise will effect on the accuracy of the program. So The higher the quality of the resolution, the lower noisy effect on the accuracy.

9.3 Future Work

We hope that this project will discuss new implementations and techniques in the area of video and image processing. Several of the areas that may be promising for further research is described in the following:

- We can use a mixture of two techniques, the HSI & RGB techniques to detect the dominant color of any sport. We can use HSI technique to detect the dominant color of the sport field, and convert H, S, I components to the corresponding RGB components to define the color range of this sport field. So by applying such a technique like this, the detection of a dominant color of any support will be easy and more adaptive, because it will depend on the color of the sport field without any restrictions.
- Porting the proposed application as a web service that can scale to large number of users. This should ease deployment, maintenance.
- Taking the advantage of using distributed system to improve the process time of our application, and use poor capabilities computers to collect their capabilities to distribute the work among them, without the need to buy a super computer to do the work if there are existing computers with low capabilities.
- Adapt the logo detection to work with more championships and different environments.
- Gathering more statistics about the played soccer game, and move from the soccer summarization area zone to the soccer analysis area zone with the use of new techniques such as object tracking.

Bibliography

[1] Ahmed.Ekin. (2003). *Sports Video Processing for Description.* Rochester, New York.

[2] R. C. Gonzalez and R. E. Wood. (1992). *Digital image processing.* Addison-Wesley, Boston,MA.

[3] K. N. Plataniotis and A. N. Venetsanopoulos. (2000). *Color image processing and applications.* Springer-Verlag, Berlin, Germany.

[4] Dian W. Tjondronegoro, BIT (1st Hons). *Content-based Video Indexing for Sports Applications.*

[5] Lei Wangy, Xu Liuy, Steve Linz, Guangyou Xuy, Heung-Yeung Shumz. *GENERIC SLOW-MOTION REPLAY DETECTION IN SPORTS VIDEO.*

[6] Chung-Lin Huang, Senior Member, IEEE, Huang-Chia Shih, Student Member, IEEE, and Chung-Yuan Chao. (August, 2006). *Semantic Analysis of Soccer Video Using Dynamic Bayesian Network.*

[7] Kongwah WAN, Xin YAN, Xinguo YU, Changsheng XU. (2003). *Real-time Goal-Mouth Detection in MPEG Soccer Video.* Singapore.

[8] http://www.soccerperformance.org/specialtopics/companalsystems.htm. (Jun 2008). *COMPUTER TECHNOLOGY IN SOCCER.*

[9] Fatih Porikli. (2002). *AUTOMATIC THRESHOLD DETERMINATION OF CENTROID-LINKAGE REGION GROWING BY MPEG-7 DOMINANT COLOR DESCRIPTORS.* USA.

[10] Matko Šarić,Hrvoje Dujmić and Domagoj Baričević. *Shot Boundary Detection in Soccer Video using Twin-comparison Algorithm and Dominant Color Region.* FESB, University of Split,Ruđera Boškovića b.b.,Split.

[11] Dian Tjondronegoro, Yi-Ping Phoebe Chen and Binh Pham. *A Statistical-driven Approach for Automatic Classification of Events in AFL Video Highlights.* Deakin University.

[12] MuratTekalp and A.Ekin. *Shot type classification by dominant color for sports video segmentation and summarization.*

[13] Shu-Ching Chen, Min Chen, Chengcui Zhang and Mei-Ling Shyu. *Exciting Event Detection Using Multi-level Multimodal Descriptors and Data Classification.* USA.

[14] Lifang Gu, Don Bone, and Graham Reynolds. *Replay Detection in Sports Video Sequences.* Australia.

[15] Reede Ren and Joemon M. Jose. *Football Video Segmentation Based on Video Production Strategy.* UK.

[16] SUN Xing-hua, YANG Jing-yu. *Inference and retrieval of soccer event.* China.

[17] Zhao Zhao, Shuqiang Jiang, Qingming Huang and Qixiang Ye. *Highlight Summarization in Soccer Video based on Goalmouth Detection.* China.

[18] M. Bertini, A. Del Bimbo, C. Torniai, C. Grana, R. Vezzani, R. Cucchiara. (2007). *Sports Video Annotation Using Enhanced HSV Histograms in Multimedia Ontologies.* Italy.